GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:21:02 ; Search time 17 Seconds
(without alignments)
486.327 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	459	100.0	1065	2	A57410	transcription fact
2	430	93.7	1075	2	A57377	transcription fact
3	114	24.8	716	2	S45262	NF-AT component -
4	96.5	21.0	718	2	JC5805	transcription fact
5	81.5	17.8	329	2	G86253	hypothetical prote
6	73.5	16.0	3488	2	T34418	hypothetical prote
7	72	15.7	2130	2	AB0821	probable exported
8	71.5	15.6	792	2	JC7122	protein kinase (EC
9	70	15.3	643	2	S54395	permease-like prot
10	70	15.3	649	1	B37953	transcription regu
11	70	15.3	652	2	G95177	hypothetical prote
12	70	15.3	652	2	E98044	hypothetical prote
13	70	15.3	660	2	G97912	hypothetical prote
14	70	15.3	661	2	E95042	hypothetical prote
15	69.5	15.1	638	2	B35816	transcription regu
16	69	15.0	649	2	A35816	transcription regu
17	67.5	14.7	314	2	I38864	E1B 19K/Bcl-2-inte
18	67	14.6	177	2	H90631	probable fimbrial
19	67	14.6	177	2	G85482	probable type-1 fi
20	67	14.6	1112	2	T32733	AMPA glutamate rec
21	66.5	14.5	1039	2	T28905	hypothetical prote
22	66	14.4	728	1	A60185	hepatocyte growth
23	66	14.4	1944	2	A59438	KIAA1424 protein [
24	65.5	14.3	652	2	S47979	aliB protein precu
25	65	14.2	635	2	T20587	hypothetical prote
26	64.5	14.1	421	2	I49734	HNF-3/fork-head ho
27	64	13.9	84	2	S76062	hypothetical prote
28	64	13.9	616	2	S38060	carboxylic acid tr
29	64	13.9	728	1	A35644	hepatocyte growth
30	63	13.7	498	2	T09436	gag polyprotein -
31	63	13.7	520	2	S47142	mating type A prot
32	63	13.7	570	2	T49181	cyclophylin-like p
33	63	13.7	1216	2	AH1535	pyruvate-flavodoxi
34	62.5	13.6	474	2	T39587	serine/threonine
35	62.5	13.6	671	2	S61693	probable membrane
36	62.5	13.6	1530	2	E82085	glutamate synthase
37	62	13.5	294	2	G97484	dihydrodipicolinat
38	62	13.5	294	2	AG2702	dihydrodipicolinat
39	62	13.5	579	2	AF2094	hypothetical prote
40	62	13.5	963	2	T40290	hypothetical prote
41	62	13.5	1127	2	T30334	AND-1 protein - Af
42	62	13.5	1856	2	C95008	immunoglobulin A1
43	61.5	13.4	279	2	C75491	probable serine es
44	61.5	13.4	441	2	B86252	hypothetical prote
45	61.5	13.4	838	2	A96557	probable receptor

ALIGNMENTS

RESULT 1
A57410
transcription factor NFATc3 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 08-Feb-1996
C;Accession: A57410
R;Ho, S.N.; Thomas, D.J.; Timmerman, L.A.; Li, X.; Francke, U.; Crabtree, G.R.
J. Biol. Chem. 270, 19898-19907, 1995
A;Title: NFATc3, a lymphoid-specific NFATc family member that is calcium-regulated an
A;Reference number: A57410; MUID:95378239; PMID:7650004
A;Accession: A57410
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1065 <HOA>
A;Cross-references: GB:U28807

Query Match 100.0%; Score 459; DB 2; Length 1065;
Best Local Similarity 100.0%; Pred. No. 5.8e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 370

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||||||||||||||||||||||
Db     371 YPLKKDSSGDQFLSVPSPFTWSKPKP 396
```

RESULT 2
A57377
transcription factor NFATx - human
C;Species: Homo sapiens (man)
C;Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 05-Nov-1999
C;Accession: A57377
R;Masuda, E.S.; Naito, Y.; Tokumitsu, H.; Campbell, D.; Saito, F.; Hannum, C.; Arai,
Mol. Cell. Biol. 15, 2697-2706, 1995
A;Title: NFATx, a novel member of the nuclear factor of activated T cells family that
A;Reference number: A57377; MUID:95257951; PMID:7739550
A;Accession: A57377
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1075 <MAS>
A;Cross-references: GB:U14510; NID:g780373; PIDN:AAA86308.1; PID:g780374
C;Keywords: transcription factor

Query Match 93.7%; Score 430; DB 2; Length 1075;
Best Local Similarity 91.9%; Pred. No. 8.7e-39;
Matches 79; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || |||||||||||||||||||||||||||||||:|||||:|||:||||:||||||||
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:20:32 ; Search time 13 Seconds
(without alignments)
274.382 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	459	100.0	1075	1	NFC3_MOUSE	P97305 mus musculu
2	430	93.7	1075	1	NFC3_HUMAN	Q12968 homo sapien
3	122	26.6	902	1	NFC4_HUMAN	Q14934 homo sapien
4	114	24.8	943	1	NFC1_HUMAN	O95644 homo sapien
5	102	22.2	822	1	NFC1_PIG	O77638 sus scrofa
6	96.5	21.0	717	1	NFC1_MOUSE	O88942 mus musculu
7	71.5	15.6	513	1	DMP1_HUMAN	Q13316 homo sapien
8	70	15.3	649	1	TFE2_MESAU	P98180 mesocricetu
9	70	15.3	652	1	ALIB_STRPN	Q51933 streptococc
10	70	15.3	660	1	ALIA_STRPN	P35592 streptococc
11	69	15.0	649	1	TFE2_RAT	P21677 rattus norv
12	67.5	14.7	314	1	NIP2_HUMAN	Q12982 homo sapien
13	67	14.6	558	1	VNNL_DROME	Q9nfp1 drosophila
14	66.5	14.5	486	1	RIK3_MOUSE	Q9qz10 mus musculu
15	66	14.4	728	1	HGF_MOUSE	Q08048 mus musculu
16	65.5	14.3	326	1	NIP2_MOUSE	O54940 mus musculu
17	65	14.2	898	1	CIZ1_HUMAN	Q9ulv3 homo sapien
18	64.5	14.1	421	1	FXJ1_MOUSE	Q61660 mus musculu
19	64.5	14.1	535	1	TX21_HUMAN	Q9ul17 homo sapien
20	64.5	14.1	708	1	ICAL_HUMAN	P20810 homo sapien
21	64	13.9	616	1	JEN1_YEAST	P36035 saccharomyc
22	64	13.9	728	1	HGF_RAT	P17945 rattus norv
23	62	13.5	500	1	GAG_HV1RH	P05890 human immun
24	62	13.5	1815	1	SHK3_RAT	Q9jlu4 rattus norv
25	61.5	13.4	367	1	IHA_HORSE	P55101 equus cabal
26	61.5	13.4	2032	1	CTOG_HUMAN	Q14008 homo sapien
27	61	13.3	508	1	GAG_SIVCZ	P17282 chimpanzee
28	61	13.3	649	1	FAF1_MOUSE	P54731 mus musculu
29	61	13.3	1203	1	MGR5_RAT	P31424 rattus norv
30	60.5	13.2	421	1	FXJ1_RAT	Q63247 rattus norv
31	60.5	13.2	503	1	CP39_RAT	P51538 rattus norv
32	60.5	13.2	586	1	HOL1_YEAST	P53389 saccharomyc
33	60.5	13.2	642	1	NA95_MOUSE	Q9r017 mus musculu
34	60.5	13.2	653	1	EGLN_MOUSE	Q63961 mus musculu
35	60.5	13.2	2895	1	HYD_DROME	P51592 drosophila
36	60	13.1	488	1	HH1R_MOUSE	P70174 mus musculu
37	60	13.1	498	1	GAG_HV1OY	P20889 human immun
38	60	13.1	529	1	YB89_YEAST	P38140 saccharomyc
39	60	13.1	607	1	RPB1_HALN1	P15351 halobacteri
40	60	13.1	611	1	HBS1_YEAST	P32769 saccharomyc
41	60	13.1	646	1	GYRB_MYCGA	P47720 mycoplasma
42	60	13.1	1069	1	ACAA_ARATH	Q9szr1 arabidopsis
43	60	13.1	1093	1	SM5B_MOUSE	Q60519 mus musculu
44	60	13.1	1856	1	GBF1_CRIGR	Q9rld7 cricetulus
45	59.5	13.0	467	1	YAIU_ECOLI	P75700 escherichia

ALIGNMENTS

RESULT 1
NFC3_MOUSE
ID NFC3_MOUSE STANDARD; PRT; 1075 AA.
AC P97305; Q60896;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear factor of activated T-cells, cytoplasmic 3 (T cell
DE transcription factor NFAT4) (NF-ATc3) (NF-AT4) (NFATx).
GN NFATC3 OR NFAT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM X1).
RC TISSUE=Thymus;
RX MEDLINE=95378239; PubMed=7650004;
RA Ho S.N., Thomas D.J., Timmerman L.A., Li X., Francke U.,
RA Crabtree G.R.;
RT "NFATc3, a lymphoid-specific NFATc family member that is
RT calcium-regulated and exhibits distinct DNA binding specificity.";
RL J. Biol. Chem. 270:19898-19907(1995).
RN [2]
RP SEQUENCE OF 12-1075 FROM N.A. (ISOFORMS X1; X2 AND DELTA-X).
RC TISSUE=Thymic lymphoma;
RX MEDLINE=97170074; PubMed=9017603;
RA Liu J., Koyano-Nakagawa N., Amasaki Y., Saito-Oha
RA Imai S.-I., Takano T., Arai N., Yokota T., Arai I
RT "Calcineurin-dependent nuclear translocation of
RT factor NFATx: molecular cloning and functional characterization...
RL Mol. Biol. Cell 8:157-170(1997).
RN [3]
RP REVIEW.
RX MEDLINE=99189746; PubMed=10089876;
RA Crabtree G.R.;
RT "Generic signals and specific outcomes: signaling through Ca2+,
RT calcineurin, and NF-AT.";
RL Cell 96:611-614(1999).
CC -!- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
CC GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2 (BY
CC SIMILARITY).
CC -!- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
CC THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
CC CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
CC NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
CC ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
CC THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
CC NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
CC MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
CC TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
CC AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
CC NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; X1 (SHOWN HERE), X2 AND DELTA-X;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.

SNC

```
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS. WEAKLY EXPRESSED IN
CC       MUSCLE, SPLEEN AND KIDNEY. ALSO EXPRESSED IN LYMPH NODE.
CC   -!- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
CC       COOPERATIVE INTERACTIONS WITH AP1 FACTORS (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
CC       CALCINEURIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D85612; BAA12833.1; -.
DR   EMBL; U28807; AAA93249.1; -.
DR   MGD; MGI:103296; Nfatc3.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; FALSE_NEG.
DR   PROSITE; PS50254; REL_2; 1.
KW   Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW   Alternative splicing; Repeat; Phosphorylation.
FT   DOMAIN       24     29       POLY-PRO.
FT   DOMAIN      109    114       CALCINEURIN-BINDING.
FT   DOMAIN      207    308       3 X SP REPEATS.
FT   REPEAT      207    223       1.
FT   REPEAT      236    252       2.
FT   REPEAT      292    308       3.
FT   DOMAIN      273    275       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      444    451       DNA-BINDING.
FT   DOMAIN      686    688       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN     1031   1040       NUCLEAR EXPORT SIGNAL.
FT   VARSPLIC    468    497       MISSING (IN ISOFORM DELTA-X).
FT   VARSPLIC   1035   1075       VNEIIGRDMSQISVSQATEVMRDTPLPGPASPDLMTSHSAH
FT                                -> DQLISDLEHQPSGSTEKWSNHSEFSCPVPFWRI (IN
FT                                ISOFORM X2).
FT   CONFLICT     12     34       LDFKLVFGEDGAPAPPPPGSRPA -> PRRVLFSVSAQLPS
FT                                RTRPGPSDL (IN REF. 2).
FT   CONFLICT     57     61       TTPLC -> NSSLG (IN REF. 2).
FT   CONFLICT     68     68       P -> LQ (IN REF. 2).
FT   CONFLICT     82     83       SH -> GY (IN REF. 2).
FT   CONFLICT     90     93       CEIP -> GDIS (IN REF. 2).
FT   CONFLICT    113    113       F -> I (IN REF. 2).
FT   CONFLICT    121    124       QQEL -> HQGT (IN REF. 2).
FT   CONFLICT    132    132       Q -> H (IN REF. 2).
FT   CONFLICT    140    140       F -> Y (IN REF. 2).
FT   CONFLICT    641    641       G -> R (IN REF. 2).
FT   CONFLICT    646    646       V -> A (IN REF. 2).
FT   CONFLICT    707    734       REDTDLPSVPSLPVPHSAQAQRPSSETG -> KEKTQICLQ
FT                                FHLCLCLILPSPEALLRDR (IN REF. 2).
FT   CONFLICT    746    746       G -> S (IN REF. 2).
FT   CONFLICT   1075   1075       H -> Q (IN REF. 2).
SQ   SEQUENCE   1075 AA;  115450 MW;  4ED38C9AA6F452BB CRC64;

  Query Match           100.0%;  Score 459;  DB 1;  Length 1075;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-41;
  Matches   86;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 380

Qy     61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
          ||||||||||||||||||||||||||
Db    381 YPLKKDSSGDQFLSVPSPFTWSKPKP 406
```

```
RESULT 2
NFC3_HUMAN
ID   NFC3_HUMAN     STANDARD;      PRT;  1075 AA.
AC   Q12968; Q14516; Q99840; Q99841; Q99842; O75211;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear factor of activated T-cells, cytoplasmic 3 (T cell
DE   transcription factor NFAT4) (NF-ATc3) (NF-AT4) (NFATx).
GN   NFATC3 OR NFAT4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B AND X2/C).
RC   TISSUE=T-cell, and Skeletal muscle;
RX   MEDLINE=95269130; PubMed=7749981;
RA   Hoey T., Sun Y.-L., Williamson K., Xu X.;
RT   "Isolation of two new members of the NF-AT gene family and functional
RT   characterization of the NF-AT proteins.";
RL   Immunity 2:461-472(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM X1).
RC   TISSUE=T-cell;
RX   MEDLINE=95257951; PubMed=7739550;
RA   Masuda E.S., Naito Y., Tokumitsu H., Campbell D., Saito F., Hannum C.,
RA   Arai K.-I., Arai N.;
RT   "NFATx, a novel member of the nuclear factor of activated T cells
RT   family that is expressed predominantly in the thymus.";
RL   Mol. Cell. Biol. 15:2697-2706(1995).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS X2; X3 AND X4).
RC   TISSUE=T-cell, and Fibroblast;
RX   MEDLINE=98430664; PubMed=9759864;
RA   Imamura R., Masuda E.S., Naito Y., Imai S.-I., Fujino T., Takano T.,
RA   Arai K.-I., Arai N.;
RT   "Carboxy-terminal 15 amino acids sequence of NFATx1 possibly created
RT   by tissue specific splicing is conserved among NFAT family proteins
RT   and is essential for transactivation activity in T cells.";
RL   J. Immunol. 161:3455-3463(1998).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM X4).
RX   MEDLINE=99425270; PubMed=10493829;
RA   Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA   Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA   Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA   Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT   "Genome duplications and other features in 12 Mb of DNA sequence from
RT   human chromosome 16p and 16q.";
RL   Genomics 60:295-308(1999).
RN   [5]
RP   MUTAGENESIS.
RX   MEDLINE=98292182; PubMed=9630228;
RA   Zhu J., Shibasaki F., Price R., Guillemot J.-C., Yano T., Doetsch V.,
RA   Wagner G., Ferrara P., McKeon F.;
RT   "Intramolecular masking of nuclear import signal on NF-AT4 by casein
RT   kinase I and MEKK1.";
RL   Cell 93:851-861(1998).
RN   [6]
RP   REVIEW.
RX   MEDLINE=99189746; PubMed=10089876;
RA   Crabtree G.R.;
RT   "Generic signals and specific outcomes: signaling through Ca2+,
RT   calcineurin, and NF-AT.";
RL   Cell 96:611-614(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
CC       GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2.
CC   -!- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
CC       THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
CC       CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
CC       NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
CC       ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
```

```
CC       THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
CC       NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
CC       MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
CC       TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
CC       AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
CC       NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.
CC   -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; X1 (SHOWN HERE), X2/C, X3, X4, A
CC       AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ISOFORM X1 IS PREDOMINANTLY EXPRESSED IN
CC       THYMUS AND IS ALSO FOUND IN PERIPHERAL BLOOD LEUKOCYTES AND
CC       KIDNEY. ISOFORM X2 IS PREDOMINANTLY EXPRESSED IN SKELETAL MUSCLE
CC       AND IS ALSO FOUND IN THYMUS, KIDNEY, TESTIS, SPLEEN, PROSTATE,
CC       OVARY, SMALL INTESTINE, HEART, PLACENTA AND PANCREAS. ISOFORM X3
CC       IS EXPRESSED IN THYMUS AND KIDNEY. ISOFORM X4 IS EXPRESSED IN
CC       THYMUS AND SKELETAL MUSCLE.
CC   -!- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
CC       COOPERATIVE INTERACTIONS WITH AP1 FACTORS.
CC   -!- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
CC       CALCINEURIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L41067; AAA79174.1; -.
DR   EMBL; U14510; AAA86308.1; -.
DR   EMBL; U85428; AAB46595.1; -.
DR   EMBL; U85429; AAB46596.1; -.
DR   EMBL; U85430; AAB46597.1; -.
DR   EMBL; AC004531; AAC27434.1; -.
DR   Genew; HGNC:7777; NFATC3.
DR   MIM; 602698; -.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; FALSE_NEG.
DR   PROSITE; PS50254; REL_2; 1.
KW   Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW   Alternative splicing; Repeat; Phosphorylation.
FT   DOMAIN       24     29       POLY-PRO.
FT   DOMAIN      109    114       CALCINEURIN-BINDING.
FT   DOMAIN      207    308       3 X SP REPEATS.
FT   REPEAT      207    223       1.
FT   REPEAT      236    252       2.
FT   REPEAT      292    308       3.
FT   DOMAIN      273    275       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      444    451       DNA-BINDING.
FT   DOMAIN      686    688       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN     1032   1041       NUCLEAR EXPORT SIGNAL.
FT   VARSPLIC    700   1075       VLMKQEHREEIDLSSVPSLPVPHPAQTQRPSSDSGCSHDSV
FT                                LSGQRSLICSIPQTYASMVTSSHLPQLQCRDESVSKEQHMI
FT                                PSPIVHQPFQVTPTPPVGSSYQPMQTNVVYNGPTCLPINAA
FT                                SSQEFDSVLFQQDATLSGLVNLGCQPLSSIPFHSSNSGSTG
FT                                HLLAHTPHSVHTLPHLQSMGYHCSNTGQRSLSSPVADQITG
FT                                QPSSQLQPITYGPSHSGSATTASPAASHPLASSPLSGPPSP
FT                                QLQPMPYQSPSSGTASSPSPATRMHSGQHSTQAQSTGQGGL
FT                                SAPSSLICHSLCDPASFPPDGATVSIKPEPEDREPNFATIG
FT                                LQDITLDDVNEIIGRDMSQISVSQGAGVSRQAPLPSPESLD
FT                                LGRSDGL -> GTRSHDGLL (IN ISOFORM A).
FT   VARSPLIC    717   1075       SLPVPHPAQTQRPSSDSGCSHDSVLSGQRSLICSIPQTYAS
FT                                MVTSSHLPQLQCRDESVSKEQHMIPSPIVHQPFQVTPTPPV
FT                                GSSYQPMQTNVVYNGPTCLPINAASSQEFDSVLFQQDATLS
FT                                GLVNLGCQPLSSIPFHSSNSGSTGHLLAHTPHSVHTLPHLQ
FT                                SMGYHCSNTGQRSLSSPVADQITGQPSSQLQPITYGPSHSG
FT                                SATTASPAASHPLASSPLSGPPSPQLQPMPYQSPSSGTASS
FT                                PSPATRMHSGQHSTQAQSTGQGGLSAPSSLICHSLCDPASF
FT                                PPDGATVSIKPEPEDREPNFATIGLQDITLDDVNEIIGRDM
FT                                SQISVSQGAGVSRQAPLPSPESLDLGRSDGL -> TLPQTS
FT                                RQTLLGSQPPSASPPTV (IN ISOFORM B).
FT   VARSPLIC   1036   1075       VNEIIGRDMSQISVSQGAGVSRQAPLPSPESLDLGRSDGL
FT                                -> DQFISDLEHQPSGSAEKWPNHSVLSCPAPFWRI (IN
FT                                ISOFORM X2/C).
FT   VARSPLIC   1036   1075       VNEIIGRDMSQISVSQGAGVSRQAPLPSPESLDLGRSDGL
FT                                -> DLFTSNNFDLLQLRPTFWPVPAGRYLRNLE (IN
FT                                ISOFORM X3).
FT   VARSPLIC   1036   1075       VNEIIGRDMSQISVSQGAGVSRQAPLPSPESLDLGRSDGL
FT                                -> GKFISDMFLK (IN ISOFORM X4).
FT   CONFLICT    702    702       M -> L (IN REF. 3).
FT   CONFLICT    831    831       L -> W (IN REF. 3).
FT   CONFLICT    899    899       A -> G (IN REF. 3).
FT   CONFLICT    923    923       A -> V (IN REF. 3).
FT   CONFLICT    935    935       A -> G (IN REF. 3).
FT   CONFLICT    947    947       L -> F (IN REF. 3).
FT   CONFLICT    960    960       A -> G (IN REF. 3).
SQ   SEQUENCE   1075 AA;  115593 MW;  0B25D61A2B56D898 CRC64;

  Query Match           93.7%;  Score 430;  DB 1;  Length 1075;
  Best Local Similarity   91.9%;  Pred. No. 2.7e-38;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || |||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
Db     381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 3
NFC4_HUMAN
ID   NFC4_HUMAN     STANDARD;      PRT;   902 AA.
AC   Q14934;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear factor of activated T-cells, cytoplasmic 4 (T cell
DE   transcription factor NFAT3) (NF-ATc4) (NF-AT3).
GN   NFATC4 OR NFAT3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=95269130; PubMed=7749981;
RA   Hoey T., Sun Y.-L., Williamson K., Xu X.;
RT   "Isolation of two new members of the NF-AT gene family and functional
RT   characterization of the NF-AT proteins.";
RL   Immunity 2:461-472(1995).
RN   [2]
RP   REVIEW.
RX   MEDLINE=99189746; PubMed=10089876;
RA   Crabtree G.R.;
RT   "Generic signals and specific outcomes: signaling through Ca2+,
RT   calcineurin, and NF-AT.";
RL   Cell 96:611-614(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
CC       GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2 AND IL-
CC       4 (BY SIMILARITY).
CC   -!- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
CC       THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
CC       CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
CC       NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
CC       ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
CC       THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
```

```
CC       NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
CC       MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
CC       TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
CC       AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
CC       NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PLACENTA, LUNG, KIDNEY,
CC       TESTIS AND OVARY. WEAKLY EXPRESSED IN SPLEEN AND THYMUS. NOT
CC       EXPRESSED IN PERIPHERAL BLOOD LYMPHOCYTES.
CC   -!- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
CC       COOPERATIVE INTERACTIONS WITH AP1 FACTORS (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
CC       CALCINEURIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L41066; AAA79175.1; -.
DR   TRANSFAC; T02462; -.
DR   Genew; HGNC:7778; NFATC4.
DR   MIM; 602699; -.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; FALSE_NEG.
DR   PROSITE; PS50254; REL_2; 1.
KW   Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW   Repeat; Phosphorylation.
FT   DOMAIN       62     69       POLY-PRO.
FT   DOMAIN      114    119       CALCINEURIN-BINDING.
FT   DOMAIN      213    293       2 APPROXIMATE SP REPEATS.
FT   REPEAT      213    229       SP 1.
FT   REPEAT      277    293       SP 2 (APPROXIMATE).
FT   DOMAIN      297    304       POLY-PRO.
FT   DOMAIN      268    270       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      430    437       DNA-BINDING.
FT   DOMAIN      672    674       NUCLEAR LOCALIZATION SIGNAL.
SQ   SEQUENCE   902 AA;  95472 MW;  E59F15F7647A47C6 CRC64;

  Query Match             26.6%;  Score 122;  DB 1;  Length 902;
  Best Local Similarity   40.0%;  Pred. No. 2.6e-05;
  Matches   34;  Conservative   11;  Mismatches   34;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :      : || |||:|| :||  ||   |  | : | |    | ||    ||
Db     311 SPGPFDYVGAPPAESIPQKTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 368

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             :|: :|  :|:|||| ||| :
Db     369 ---RKEVAGMDYLAVPSPLAWSKAR 390


RESULT 4
NFC1_HUMAN
ID   NFC1_HUMAN     STANDARD;      PRT;   943 AA.
AC   O95644; Q12865; Q15793;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nuclear factor of activated T-cells, cytoplasmic 1 (NFAT transcription
DE   complex cytosolic component) (NF-ATc1) (NF-ATc).
GN   NFATC1 OR NFATC OR NFAT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM A-ALPHA).
RC   TISSUE=T-cell, and Peripheral blood lymphocytes;
RX   MEDLINE=94261186; PubMed=8202141;
RA   Northrop J.P., Ho S.N., Chen L., Thomas D.J., Timmerman L.A.,
RA   Nolan G.P., Admon A., Crabtree G.R.;
RT   "NF-AT components define a family of transcription factors targeted in
RT   T-cell activation.";
RL   Nature 369:497-502(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM B-BETA).
RC   TISSUE=B-cell;
RX   MEDLINE=96355439; PubMed=8702849;
RA   Park J., Takeuchi A., Sharma S.;
RT   "Characterization of a new isoform of the NFAT (nuclear factor of
RT   activated T cells) gene family member NFATc.";
RL   J. Biol. Chem. 271:20914-20921(1996).
RN   [3]
RP   ERRATUM.
RA   Park J., Takeuchi A., Sharma S.;
RL   J. Biol. Chem. 271:33705-33705(1996).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS A-ALPHA; B-ALPHA AND C-BETA).
RC   TISSUE=B-cell lymphoma;
RX   MEDLINE=99170294; PubMed=10072078;
RA   Chuvpilo S., Zimmer M., Kerstan A., Gloeckner J., Avots A., Escher C.,
RA   Fischer C., Inashkina I., Jankevics E., Berberich-Siebelt F.,
RA   Schmitt E., Serfling E.;
RT   "Alternative polyadenylation events contribute to the induction of
RT   NF-ATc in effector T cells.";
RL   Immunity 10:261-269(1999).
RN   [5]
RP   MUTAGENESIS.
RX   MEDLINE=20119316; PubMed=10652349;
RA   Porter C.M., Havens M.A., Clipstone N.A.;
RT   "Identification of amino acid residues and protein kinases involved in
RT   the regulation of NFATc subcellular localization.";
RL   J. Biol. Chem. 275:3543-3551(2000).
RN   [6]
RP   ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX   MEDLINE=99288090; PubMed=10358178;
RA   Chuvpilo S., Avots A., Berberich-Siebelt F., Gloeckner J., Fischer C.,
RA   Kerstan A., Escher C., Inashkina I., Hlubek F., Jankevics E.,
RA   Brabletz T., Serfling E.;
RT   "Multiple NF-ATc isoforms with individual transcriptional properties
RT   are synthesized in T lymphocytes.";
RL   J. Immunol. 162:7294-7301(1999).
RN   [7]
RP   REVIEW.
RX   MEDLINE=99189746; PubMed=10089876;
RA   Crabtree G.R.;
RT   "Generic signals and specific outcomes: signaling through Ca2+,
RT   calcineurin, and NF-AT.";
RL   Cell 96:611-614(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
CC       GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2 OR IL-4
CC       GENE TRANSCRIPTION. ALSO CONTROL GENE EXPRESSION IN EMBRYONIC
CC       CARDIAC CELLS. COULD REGULATE NOT ONLY THE ACTIVATION AND
CC       PROLIFERATION BUT ALSO THE DIFFERENTIATION AND PROGRAMMED DEATH OF
CC       T-LYMPHOCYTES AS WELL AS LYMPHOID AND NONLYMPHOID CELLS.
CC   -!- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
CC       THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
CC       CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
CC       NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
CC       ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
CC       THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
CC       NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
CC       MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
CC       TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
CC       AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
CC       NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; A-ALPHA, A-BETA, B-
CC       ALPHA, B-BETA, C-ALPHA (SHOWN HERE) AND C-BETA; ARE PRODUCED BY
```

```
CC       ALTERNATIVE SPLICING. THERE MIGHT BE ADDITIONAL ISOFORMS PRODUCED
CC       BY ALTERNATIVE INITIATION AT MET-37 OF ISOFORMS ALPHA. NFATC/C IS
CC       THE STRONGEST ACTIVATOR OF GENE TRANSCRIPTION, FOLLOWED BY
CC       NFATC/A, WHEREAS NFATC/B IS THE WEAKEST. NFATC/B AND NFATC/C, BOTH
CC       PRESENT IN T-CELLS CAN MODULATE THEIR TRANSCRIPTIONAL ACTIVITY.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, PERIPHERAL LEUKOCYTES AS
CC       T-CELLS AND SPLEEN. NFTAC/A IS PREFERENTIALLY EXPRESSED IN
CC       EFFECTOR T-CELLS (THYMUS AND PERIPHERAL LEUKOCYTES) WHEREAS
CC       NFATC/B AND NFATC/C ARE PREFERENTIALLY EXPRESSED IN NAIVE T-CELLS
CC       (SPLEEN). NFATC/B IS EXPRESSED IN NAIVE T-CELLS AFTER FIRST
CC       ANTIGEN EXPOSURE AND ISOFORM A IS EXPRESSED IN EFFECTOR T-CELLS
CC       AFTER SECOND ANTIGEN EXPOSURE.
CC   -!- INDUCTION: ONLY NFATC/A IS INDUCIBLY EXPRESSED IN T LYMPHOCYTES
CC       UPON ACTIVATION OF THE T-CELL RECEPTOR (TCR) COMPLEX. INDUCED
CC       AFTER CO-ADDITION OF PHORBOL 12-MYRISTATE 13-ACETATE (PMA) AND
CC       IONOMYCIN. ALSO INDUCED AFTER CO-ADDITION OF 12-O-
CC       TETRADECANOYLPHORBOL-13-ACETATE (TPA) AND IONOMYCIN. WEAKLY
CC       INDUCED WITH PMA, IONOMYCIN AND CYCLOSPORIN A.
CC   -!- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
CC       COOPERATIVE INTERACTIONS WITH AP1 FACTORS.
CC   -!- DOMAIN: THE N-TERMINAL TRANSACTIVATION DOMAIN (TAD-A) BINDS TO
CC       AND IS ACTIVATED BY CBP/P300. THE DEPHOSPHORYLATED FORM CONTAINS
CC       TWO UNMASKED NUCLEAR LOCALIZATION SIGNALS (NLS), WHICH ALLOW
CC       TRANSLOCATION OF THE PROTEIN TO THE NUCLEUS.
CC   -!- DOMAIN: NFATC/C HAS A C-TERMINAL PART WITH AN ADDITIONAL TRANS-
CC       ACTIVATION DOMAIN, TAD-B, WHICH ACTS AS A TRANSCRIPTIONAL
CC       ACTIVATOR. NFATC/B HAS A SHORTER C-TERMINAL PART WITHOUT COMPLETE
CC       TAD-B WHICH ACTS AS A TRANSCRIPTIONAL REPRESSOR.
CC   -!- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
CC       CALCINEURIN.
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08015; AAA19601.1; -.
DR   EMBL; U59736; AAC50869.1; -.
DR   EMBL; U80917; AAD00450.1; -.
DR   EMBL; U80918; AAD00451.1; -.
DR   EMBL; U80919; AAD00452.1; -.
DR   TRANSFAC; T01945; -.
DR   Genew; HGNC:7775; NFATC1.
DR   MIM; 600489; -.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; FALSE_NEG.
DR   PROSITE; PS50254; REL_2; 1.
KW   Transcription regulation; Activator; Repressor; Nuclear protein;
KW   DNA-binding; Alternative splicing; Phosphorylation; Repeat.
FT   DOMAIN      118    123       CALCINEURIN-BINDING.
FT   DOMAIN      126    218       TRANS-ACTIVATION DOMAIN A (TAD-A).
FT   DOMAIN      203    298       3 X SP REPEATS.
FT   REPEAT      203    219       1.
FT   REPEAT      233    249       2.
FT   REPEAT      282    298       3.
FT   DOMAIN      265    267       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      310    321       NUCLEAR EXPORT SIGNAL.
FT   DOMAIN      439    446       DNA-BINDING.
FT   DOMAIN      682    684       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      703    943       TRANS-ACTIVATION DOMAIN B (TAD-B).
FT   DOMAIN      924    933       NUCLEAR EXPORT SIGNAL.
FT   MOD_RES     117    117       PHOSPHORYLATION.
FT   VARSPLIC      1     42       MPSTSFPVPSKFPLGPAAAVFGRGETLGPAPRAGGTMKSAE
FT                                E -> MTGLEDQEFDFEFLFEFNQRDEGAAAAAP (IN
FT                                ISOFORM A-BETA, ISOFORM B-BETA AND
FT                                ISOFORM C-BETA).
FT   VARSPLIC    698    716       VPIIKTEPTDDYEPAPTCG -> GNAIFLTVSREHERVGCF
FT                                F (IN ISOFORM A-ALPHA AND ISOFORM A-
FT                                BETA).
FT   VARSPLIC    717    943       MISSING (IN ISOFORM A-ALPHA AND ISOFORM
FT                                A-BETA).
FT   VARSPLIC    826    943       MISSING (IN ISOFORM B-ALPHA AND ISOFORM
FT                                B-BETA).
FT   MUTAGEN     169    169       S->A: NO EFFECT ON SUBCELLULAR
FT                                LOCALIZATION.
FT   MUTAGEN     172    172       S->A: PARTIAL NUCLEAR TRANSLOCATION.
FT   MUTAGEN     187    187       S->A: NO EFFECT ON SUBCELLULAR
FT                                LOCALIZATION.
FT   CONFLICT    232    232       G -> S (IN REF. 2).
FT   CONFLICT    235    235       R -> Q (IN REF. 1).
SQ   SEQUENCE   943 AA;  101242 MW;  E72FAB10ECEB2D66 CRC64;

  Query Match             24.8%;  Score 114;  DB 1;  Length 943;
  Best Local Similarity   34.7%;  Pred. No. 0.0002;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 5
NFC1_PIG
ID   NFC1_PIG       STANDARD;      PRT;   822 AA.
AC   O77638;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nuclear factor of activated T-cells, cytoplasmic 1 (NFAT transcription
DE   complex cytosolic component) (NF-ATc1) (NF-ATc) (NFATmac).
GN   NFATC1 OR NFATC OR NFAT2.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Alveolar macrophage;
RX   MEDLINE=98342378; PubMed=9677199;
RA   Miskin J.E., Abrams C.C., Goatley L.C., Dixon L.K.;
RT   "A viral mechanism for inhibition of the cellular phosphatase
RT   calcineurin.";
RL   Science 281:562-565(1998).
CC   -!- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
CC       GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2 OR IL-4
CC       GENE TRANSCRIPTION. ALSO CONTROL GENE EXPRESSION IN EMBRYONIC
CC       CARDIAC CELLS. COULD REGULATE NOT ONLY THE ACTIVATION AND
CC       PROLIFERATION BUT ALSO THE DIFFERENTIATION AND PROGRAMMED DEATH OF
CC       T-LYMPHOCYTES AS WELL AS LYMPHOID AND NONLYMPHOID CELLS.
CC   -!- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
CC       THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
CC       CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
CC       NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
CC       ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
CC       THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
CC       NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
CC       MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
CC       TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
CC       AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
CC       NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.
CC   -!- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
CC       COOPERATIVE INTERACTIONS WITH AP1 FACTORS (BY SIMILARITY).
```

```
CC   -!- DOMAIN: THE N-TERMINAL TRANSACTVATION DOMAIN (TAD-A) BINDS TO
CC       AND IS ACTIVATED BY CBP/P300. THE DEPHOSPHORYLATED FORM CONTAINS
CC       TWO UNMASKED NUCLEAR LOCALIZATION SIGNALS (NLS), WHICH ALLOW
CC       TRANSLOCATION OF THE PROTEIN TO THE NUCLEUS (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
CC       CALCINEURIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF069996; AAC27301.2; -.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; FALSE_NEG.
DR   PROSITE; PS50254; REL_2; 1.
KW   Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW   Phosphorylation; Repeat.
FT   DOMAIN       25     32       POLY-ALA.
FT   DOMAIN      110    115       CALCINEURIN-BINDING.
FT   DOMAIN      118    210       TRANS-ACTIVATION DOMAIN A (TAD-A).
FT   DOMAIN      195    290       3 X SP REPEATS.
FT   REPEAT      195    211       1.
FT   REPEAT      225    241       2.
FT   REPEAT      274    290       3.
FT   DOMAIN      257    259       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      302    313       NUCLEAR EXPORT SIGNAL.
FT   DOMAIN      429    436       DNA-BINDING.
FT   DOMAIN      672    674       NUCLEAR LOCALIZATION SIGNAL.
FT   MOD_RES     109    109       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   822 AA;  88009 MW;  C891D851B3644833 CRC64;

  Query Match          22.2%;  Score 102;  DB 1;  Length 822;
  Best Local Similarity  32.4%;  Pred. No. 0.0032;
  Matches   23;  Conservative   20;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:| |||: | :  |  |:|  ::| |   |: :   ::    |: ::| :  ||:|||
Db     323 VPVKARKTALDHSPSLALKVEPAAEDLGATPPTSDFPPEEFPPPFQH-IRKGAFCDQYLSV 381

Qy      76 PS-PFTWSKPK 85
           |  |: |::|:
Db     382 PQHPYPWARPR 392


RESULT 6
NFC1_MOUSE
ID   NFC1_MOUSE     STANDARD;      PRT;   717 AA.
AC   O88942; O70345;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nuclear factor of activated T-cells, cytoplasmic 1 (NFAT transcription
DE   complex cytosolic component) (NF-ATc1) (NF-ATc).
GN   NFATC1 OR NFATC OR NFAT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM BETA).
RX   MEDLINE=98049829; PubMed=9388475;
RA   Pan S., Koyano-Nakagawa N., Tsuruta L., Amasaki Y., Yokota T.,
RA   Mori S., Arai N., Arai K.-I.;
RT   "Molecular cloning and functional characterization of murine cDNA
RT   encoding transcription factor NFATc.";
RL   Biochem. Biophys. Res. Commun. 240:314-323(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   STRAIN=BALB/c; TISSUE=Mast cells, and Fetal liver;
RX   MEDLINE=99172218; PubMed=10072529;
RA   Sherman M.A., Powell D.R., Weiss D.L., Brown M.A.;
RT   "NF-ATc Isoforms are differentially expressed and regulated in murine
RT   T and mast cells.";
RL   J. Immunol. 162:2820-2828(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE
CC       GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2 OR IL-4
CC       GENE TRANSCRIPTION. ALSO CONTROL GENE EXPRESSION IN EMBRYONIC
CC       CARDIAC CELLS. COULD REGULATE NOT ONLY THE ACTIVATION AND
CC       PROLIFERATION BUT ALSO THE DIFFERENTIATION AND PROGRAMMED DEATH OF
CC       T-LYMPHOCYTES AS WELL AS LYMPHOID AND NONLYMPHOID CELLS (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX
CC       THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING
CC       CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT
CC       NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE
CC       ACTIVATING PROTEIN-1 FAMILY, MAF, GATA4 AND CBP/300 CAN ALSO BIND
CC       THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND
CC       NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-
CC       MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT
CC       TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED
CC       AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF
CC       NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING. THERE MIGHT BE AN ADDITIONAL
CC       ISOFORM PRODUCED BY ALTERNATIVE INITIATION AT MET-37 OF ISOFORM
CC       ALPHA.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN SPLEEN, LUNG, SKELETAL MUSCLE,
CC       THYMUS AND SKIN. WEAKLY EXPRESSED IN HEART, BRAIN, LIVER AND
CC       KIDNEY. NOT EXPRESSED IN TESTIS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS AT E7 DAY AND INCREASES
CC       UNTIL E17 DAY. STRONGLY EXPRESSED IN THYMUS, LUNG AND
CC       SUBMANDIBULAR GLAND AND WEAKLY IN SKELETAL MUSCLE AND HEART.
CC   -!- INDUCTION: INDUCIBLY EXPRESSED IN T LYMPHOCYTES UPON ACTIVATION OF
CC       THE T-CELL RECEPTOR (TCR) COMPLEX. INDUCED AFTER ADDITION OF
CC       PHORBOL 12-MYRISTATE 13-ACETATE (PMA).
CC   -!- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND
CC       COOPERATIVE INTERACTIONS WITH AP1 FACTORS (BY SIMILARITY).
CC   -!- DOMAIN: THE N-TERMINAL TRANSACTVATION DOMAIN (TAD-A) BINDS TO
CC       AND IS ACTIVATED BY CBP/P300. THE DEPHOSPHORYLATED FORM CONTAINS
CC       TWO UNMASKED NUCLEAR LOCALIZATION SIGNALS (NLS), WHICH ALLOW
CC       TRANSLOCATION OF THE PROTEIN TO THE NUCLEUS (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED BY NFATC-KINASE; DEPHOSPHORYLATED BY
CC       CALCINEURIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF049606; AAC05505.1; -.
DR   EMBL; AF087434; AAC36725.1; -.
DR   MGD; MGI:102469; Nfatc1.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS01204; REL_1; FALSE_NEG.
DR   PROSITE; PS50254; REL_2; 1.
KW   Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW   Alternative splicing; Phosphorylation; Repeat.
```

```
FT   DOMAIN       120    125       CALCINEURIN-BINDING.
FT   DOMAIN       128    220       TRANS-ACTIVATION DOMAIN A (TAD-A).
FT   DOMAIN       205    300       3 X SP REPEATS.
FT   REPEAT       205    221       1.
FT   REPEAT       235    251       2.
FT   REPEAT       284    300       3.
FT   DOMAIN       267    269       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN       312    323       NUCLEAR EXPORT SIGNAL.
FT   DOMAIN       440    447       DNA-BINDING.
FT   DOMAIN       683    685       NUCLEAR LOCALIZATION SIGNAL.
FT   MOD_RES      119    119       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC       1     42       MPNTSFPVPSKFPLGPPAAVCGSGETLRPAPPSGGTMKAAE
FT                                 E -> MTGLEQDPEFDFDFLFEFDQSGGGAAAA (IN
FT                                 ISOFORM BETA).
SQ   SEQUENCE   717 AA;  77833 MW;  58837C6CC085268D CRC64;

  Query Match             21.0%;  Score 96.5;  DB 1;  Length 717;
  Best Local Similarity   30.4%;  Pred. No. 0.011;
  Matches   24;  Conservative   17;  Mismatches   19;  Indels   19;  Gaps    3;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYP--------LKKDS 67
           :|:|:|||: : |  :  |:|   :| |   |:          | :|        |:| :
Db     333 VPIKSRKTALEHAPSVALKVEPAGEDLGTTPPT----------SDFPPEEYTFQHLRKGA 382

Qy      68 SGDQFLSVP-SPFTWSKPK 85
             :|:|||| : : |:|||
Db     383 FCEQYLSVPQASYQWAKPK 401


RESULT 7
DMP1_HUMAN
ID   DMP1_HUMAN     STANDARD;      PRT;   513 AA.
AC   Q13316; O43265;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
DE   protein-1) (DMP-1).
GN   DMP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Molar;
RX   MEDLINE=97321043; PubMed=9177774;
RA   Hirst K.L., Simmons D., Feng J., Aplin H., Dixon M.J., McDougall M.;
RT   "Elucidation of the sequence and the genomic organization of the human
RT   dentin matrix acidic phosphoprotein 1 (DMP1) gene: exclusion of the
RT   locus from a causative role in the pathogenesis of dentinogenesis
RT   imperfecta type II.";
RL   Genomics 42:38-45(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   McDougall M., Juan X., Simmons D., Feng J.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.
RX   MEDLINE=96163890; PubMed=8586437;
RA   Aplin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;
RT   "Mapping of the human dentin matrix acidic phosphoprotein gene (DMP1)
RT   to the dentinogenesis imperfecta type II critical region at chromosome
RT   4q21.";
RL   Genomics 30:347-349(1995).
CC   -!- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF
CC       EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN
CC       ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U89012; AAC51332.1; -.
DR   EMBL; U34037; AAA97602.1; -.
DR   EMBL; U65378; AAB87728.1; -.
DR   Genew; HGNC:2932; DMP1.
DR   MIM; 600980; -.
KW   Extracellular matrix; Signal; Alternative splicing.
FT   SIGNAL         1     16       POTENTIAL.
FT   CHAIN         17    513       DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT   CARBOHYD      25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     467    467       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE         364    366       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC      46     61       MISSING (IN ISOFORM 2).
FT   CONFLICT      69     69       S -> C (IN REF. 2).
SQ   SEQUENCE   513 AA;  55782 MW;  2C1FDE319A5D106F CRC64;

  Query Match             15.6%;  Score 71.5;  DB 1;  Length 513;
  Best Local Similarity   26.0%;  Pred. No. 3.4;
  Matches   27;  Conservative   13;  Mismatches   21;  Indels   43;  Gaps    5;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGS- 59
           |: || |          ||:: ::||       :: |::| |  ||  |  ::|||
Db      37 LAQAPTP---------PLESSESSEGS--------KVSSEEQANEDPSDSTQSEEGLGSD 79

Qy      60 --QYPLK------------------KDSSGDQFL----SVPSP 78
             ||  :                  :| |||       | | |
Db      80 DHQYIYRLAGGFSRSTGKGGDDKDDDEDDSGDDTFGDDDSGPGP 123


RESULT 8
TFE2_MESAU
ID   TFE2_MESAU     STANDARD;      PRT;   649 AA.
AC   P98180;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcription factor E2-alpha (Immunoglobulin enhancer binding factor
DE   E12/E47) (Transcription factor-3) (TCF-3) (Transcription regulator
DE   Pan).
GN   TCF3 OR PAN.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Insulinoma;
RX   MEDLINE=91246228; PubMed=1710033;
RA   German M.S., Blanar M.A., Nelson C., Moss L.G., Rutter W.J.;
RT   "Two related helix-loop-helix proteins participate in separate cell-
RT   specific complexes that bind the insulin enhancer.";
RL   Mol. Endocrinol. 5:292-299(1991).
CC   -!- FUNCTION: HETERODIMERS BETWEEN TCF3 AND TISSUE-SPECIFIC BASIC
CC       HELIX-LOOP-HELIX (BHLH) PROTEINS PLAY MAJOR ROLES IN DETERMINING
CC       TISSUE-SPECIFIC CELL FATE DURING EMBRYOGENESIS, LIKE MUSCLE OR
CC       EARLY B-CELL DIFFERENTIATION. DIMERS BIND DNA ON E-BOX MOTIFS: 5'-
CC       CANNTG-3'. BINDS TO THE KAPPA-E2 SITE IN THE KAPPA IMMUNOGLOBULIN
CC       GENE ENHANCER (BY SIMILARITY).
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN (BY SIMILARITY).
```

```
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; E47/PAN-1 AND E12/PAN-2 (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: PHOSPHORYLATED FOLLOWING NGF STIMULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS.
DR   PIR; A37953; A37953.
DR   PIR; B37953; B37953.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS00038; HLH_1; 1.
DR   PROSITE; PS50888; HLH_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Alternative splicing; Phosphorylation.
FT   DOMAIN      171    177       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      385    420       LEUCINE-ZIPPER (POTENTIAL).
FT   DNA_BIND    542    556       BASIC DOMAIN.
FT   DOMAIN      557    600       HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   VARSPLIC    528    596       DEDDLLPPEQKAEREKERRVANNARERLRVRDINEAFKELG
FT                                RMCQLHLSSEKPQTKLLILHQAVAVILS -> STDEVLSLE
FT                                EKDLRDRERRMANNARERVRVRDINEAFRELGRICQLHLKS
FT                                DKAQTKLLILQQAVQVILG (IN ISOFORM E47).
SQ   SEQUENCE   649 AA;  67322 MW;  3C35F296316EB34D CRC64;

  Query Match             15.3%;  Score 70;  DB 1;  Length 649;
  Best Local Similarity   26.1%;  Pred. No. 6.5;
  Matches   30;  Conservative    6;  Mismatches   17;  Indels   62;  Gaps    6;

Qy      27 QAAILPGKLEICSDDQGNLSPS----------------RETSVDDGLGSQ---------- 60
           ||  |||:| : |   | ||||                 |  : | || :|
Db     123 QAGFLPGELGLSS--PGPLSPSGVKSGSQYYPSYPSNPRRRAADSGLDTQSKKVRKVPPG 180

Qy      61 -----YPLKKDSSGDQF--------------LSVPSPF-----------TWSKP 84
                ||    ||||  :              : ||||           || |
Db     181 LPSSVYP---SSSGDSYGRDAAAYPSAKTPGSAYPSPFYVADGSLHPSAELWSPP 232


RESULT 9
ALIB_STRPN
ID   ALIB_STRPN     STANDARD;      PRT;   652 AA.
AC   Q51933;
DT   15-DEC-1998 (Rel. 37, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oligopeptide-binding protein aliB precursor.
GN   ALIB OR SP1527.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R800;
RX   MEDLINE=94328326; PubMed=8051706;
RA   Alloing G., de Philip P., Claverys J.-P.;
RT   "Three highly homologous membrane-bound lipoproteins participate in
RT   oligopeptide transport by the Ami system of the Gram-positive
RT   Streptococcus pneumoniae.";
RL   J. Mol. Biol. 241:44-58(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC       PROTEIN FAMILY 5.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z16082; CAA78896.1; -.
DR   EMBL; AE007448; AAK75616.1; -.
DR   TIGR; SP1527; -.
DR   InterPro; IPR000914; SBP_bac_5.
DR   Pfam; PF00496; SBP_bac_5; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR   PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW   Peptide transport; Transport; Membrane; Lipoprotein; Signal;
KW   Complete proteome.
FT   SIGNAL        1     24       PROBABLE.
FT   CHAIN        25    652       OLIGOPEPTIDE-BINDING PROTEIN ALIB.
FT   LIPID        25     25       N-ACYL DIGLYCERIDE (PROBABLE).
FT   CONFLICT     55     55       A -> R (IN REF. 1).
FT   CONFLICT     79     80       SL -> HI (IN REF. 1).
FT   CONFLICT    123    124       LQ -> FE (IN REF. 1).
FT   CONFLICT    501    501       G -> E (IN REF. 1).
SQ   SEQUENCE   652 AA;  72562 MW;  169B67FD78CFF0CF CRC64;

  Query Match             15.3%;  Score 70;  DB 1;  Length 652;
  Best Local Similarity   30.3%;  Pred. No. 6.6;
  Matches   27;  Conservative   13;  Mismatches   31;  Indels   18;  Gaps    5;

Qy       2 SPAPFPFQYCVETDIPLKTRKTSEDQAA-------ILPGKLEICSDDQGNLSPSRE---T 51
           | |   : |   :| |       :|::||       :: | ||  :|  ||: ||     |
Db      29 STASKTYNYVYSSD-PSSLNYLAENRAATSDIVANLVDGLLE--NDQYGNIIPSLAEDWT 85

Qy      52 SVDDGLGSQYPLKKD-----SSGDQFLSV 75
               |||   | |:||     | |:::  |
Db      86 VSQDGLTYTYKLRKDAKWFTSEGEEYAPV 114


RESULT 10
ALIA_STRPN
ID   ALIA_STRPN     STANDARD;      PRT;   660 AA.
AC   P35592; Q54782; O54620; O52228;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Oligopeptide-binding protein aliA precursor (Exported protein 1).
GN   ALIA OR EXP1 OR PLPA OR SP0366.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R800;
RX   MEDLINE=94328326; PubMed=8051706;
RA   Alloing G., de Philip P., Claverys J.-P.;
RT   "Three highly homologous membrane-bound lipoproteins participate in
RT   oligopeptide transport by the Ami system of the Gram-positive
RT   Streptococcus pneumoniae.";
RL   J. Mol. Biol. 241:44-58(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11906, SP-496, SP-VA92, and SP-VA96;
```

```
RX   MEDLINE=98125733; PubMed=9466257;
RA   Coffey T.J., Enright M.C., Daniels M., Morona J.K., Morona R.,
RA   Hryniewicz W., Paton J.C., Spratt B.G.;
RT   "Recombinational exchanges at the capsular polysaccharide
RT   biosynthetic locus lead to frequent serotype changes among natural
RT   isolates of Streptococcus pneumoniae.";
RL   Mol. Microbiol. 27:73-83(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
RN   [4]
RP   SEQUENCE OF 18-660 FROM N.A.
RC   STRAIN=R6X;
RX   MEDLINE=95020610; PubMed=7523829;
RA   Pearce B., Naughton A.M., Masure H.R.;
RT   "Peptide permeases modulate transformation in Streptococcus
RT   pneumoniae.";
RL   Mol. Microbiol. 12:881-892(1994).
RN   [5]
RP   SEQUENCE OF 347-509 FROM N.A.
RC   STRAIN=R6X;
RX   MEDLINE=95020625; PubMed=7934910;
RA   Pearce B.J., Yin Y.B., Masure H.R.;
RT   "Genetic identification of exported proteins in Streptococcus
RT   pneumoniae.";
RL   Mol. Microbiol. 9:1037-1050(1993).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Probable).
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC       PROTEIN FAMILY 5.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35135; CAA84507.1; -.
DR   EMBL; AE007348; AAK74534.1; ALT_INIT.
DR   EMBL; AF030359; AAC38676.1; -.
DR   EMBL; AF030360; AAC38681.1; -.
DR   EMBL; AF030361; AAC38686.1; -.
DR   EMBL; AF030364; AAC38703.1; -.
DR   EMBL; L20556; AAA26952.1; -.
DR   TIGR; SP0366; -.
DR   InterPro; IPR000914; SBP_bac_5.
DR   Pfam; PF00496; SBP_bac_5; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR   PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW   Peptide transport; Transport; Membrane; Lipoprotein; Signal;
KW   Complete proteome.
FT   SIGNAL        1     22       PROBABLE.
FT   CHAIN        23    660       OLIGOPEPTIDE-BINDING PROTEIN ALIA.
FT   LIPID        23     23       N-ACYL DIGLYCERIDE (PROBABLE).
FT   VARIANT       7      7       F -> L (IN STRAIN R800).
FT   VARIANT      27     27       G -> D (IN STRAINS NCTC 11906, SP-VA92,
FT                                SP-496 AND SP-VA96).
FT   VARIANT      30     30       T -> A (IN STRAINS R800 AND R6X).
FT   VARIANT     166    166       E -> D (IN STRAINS NCTC 11906, SP-VA92,
FT                                SP-496 AND SP-VA96).
FT   VARIANT     246    246       V -> I (IN STRAINS R800 AND R6X).
FT   VARIANT     368    368       L -> I (IN STRAINS NCTC 11906, SP-VA92,
FT                                SP-496 AND SP-VA96).
FT   VARIANT     432    432       Q -> T (IN STRAINS R800, R6X, NCTC 11906,
FT                                SP-VA92, SP-496 AND SP-VA96).
FT   VARIANT     559    559       T -> A (IN STRAINS NCTC 11906, SP-VA92,
FT                                SP-496 AND SP-VA96).
FT   VARIANT     612    612       T -> A (IN STRAIN SP-496).
FT   CONFLICT     18     19       TT -> GV (IN REF. 4).
FT   CONFLICT    137    137       L -> P (IN REF. 4).
FT   CONFLICT    420    420       A -> R (IN REF. 1).
SQ   SEQUENCE   660 AA;  73079 MW;  702556F92EC055E8 CRC64;

  Query Match             15.3%;  Score 70;  DB 1;  Length 660;
  Best Local Similarity   34.1%;  Pred. No. 6.7;
  Matches   28;  Conservative   12;  Mismatches   26;  Indels   16;  Gaps    6;

Qy       8 FQYCVETDIP----LKTRK--TSEDQAAILPGKLEICSDDQGNLSPS--RETSVD-DGLG 58
           | |  |||      | | |  |:   : :: | ||  :|  ||  ||   : ||  ||| 
Db      36 FSYIYETDPDNLNYLTTAKAATANITSNVVDGLLE--NDRYGNFVPSMAEDWSVSKDGLT 93

Qy      59 SQYPLKKD-----SSGDQFLSV 75
             | ::||     | |::: :|
Db      94 YTYTIRKDAKWYTSEGEEYAAV 115


RESULT 11
TFE2_RAT
ID   TFE2_RAT       STANDARD;      PRT;   649 AA.
AC   P21677; P21676; Q08440;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcription factor E2-alpha (Immunoglobulin enhancer binding factor
DE   E12/E47) (Transcription factor-3) (TCF-3) (Transcription regulator
DE   Pan).
GN   TCF3 OR PAN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS E12 AND E47).
RX   MEDLINE=90346284; PubMed=2200736;
RA   Nelson C., Shen L.-P., Meister A., Fodor E., Rutter W.J.;
RT   "Pan: a transcriptional regulator that binds chymotrypsin, insulin,
RT   and AP-4 enhancer motifs.";
RL   Genes Dev. 4:1035-1043(1990).
RN   [2]
RP   SEQUENCE OF 427-649 FROM N.A. (ISOFORM E12).
RX   MEDLINE=92115315; PubMed=1766666;
RA   Metz R., Ziff E.;
RT   "The helix-loop-helix protein rE12 and the C/EBP-related factor rNFIL-
RT   6 bind to neighboring sites within the c-fos serum response element.";
RL   Oncogene 6:2165-2178(1991).
CC   -!- FUNCTION: HETERODIMERS BETWEEN TCF3 AND TISSUE-SPECIFIC BASIC
CC       HELIX-LOOP-HELIX (BHLH) PROTEINS PLAY MAJOR ROLES IN DETERMINING
CC       TISSUE-SPECIFIC CELL FATE DURING EMBRYOGENESIS, LIKE MUSCLE OR
CC       EARLY B-CELL DIFFERENTIATION. DIMERS BIND DNA ON E-BOX MOTIFS: 5'-
CC       CANNTG-3'. BINDS TO THE KAPPA-E2 SITE IN THE KAPPA IMMUNOGLOBULIN
CC       GENE ENHANCER (BY SIMILARITY). BINDS TO THE CONSENSUS
CC       SEQUENCE CAC/GCTGT/C PRESENT, IN THE CHYMOTRYPSIN, INSULIN, AP-4,
CC       AND SEVERAL OTHER GENE ENHANCER MOTIFS.
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; E47/PAN-1 AND E12/PAN-2 (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: PHOSPHORYLATED FOLLOWING NGF STIMULATION.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
```

```
CC       TRANSCRIPTION FACTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54549; CAA38421.1; -.
DR   EMBL; X62323; CAA44199.1; -.
DR   EMBL; S77532; AAB21103.1; -.
DR   PIR; A35816; A35816.
DR   PIR; B35816; B35816.
DR   TRANSFAC; T00674; -.
DR   TRANSFAC; T00675; -.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS00038; HLH_1; 1.
DR   PROSITE; PS50888; HLH_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Alternative splicing; Phosphorylation.
FT   DOMAIN      171    177       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      385    420       LEUCINE-ZIPPER (POTENTIAL).
FT   DNA_BIND    542    556       BASIC DOMAIN.
FT   DOMAIN      557    600       HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   VARSPLIC    525    596       PDEDEDDLLPPEQKAEREKERRVANNARERLRVRDINEAFK
FT                                ELGRMCQLHLSTEKPQTKLLILHQAVAVILS -> STDEVL
FT                                SLEEKDLRDRERRMANNARERVRVRDINEAFRELGRMCQLH
FT                                LKSDKAQTKLLILQQAVQVILG (IN ISOFORM E47).
FT   CONFLICT    167    167       L -> LA (IN REF. 1; CAA44199).
FT   CONFLICT    427    427       P -> A (IN REF. 2).
FT   CONFLICT    508    508       D -> DH (IN REF. 2).
FT   CONFLICT    576    577       ST -> NS (IN REF. 2).
FT   CONFLICT    637    637       P -> T (IN REF. 2).
SQ   SEQUENCE   649 AA;  67654 MW;  882F19EDB47D14EA CRC64;

  Query Match             15.0%;  Score 69;  DB 1;  Length 649;
  Best Local Similarity   26.1%;  Pred. No. 8.4;
  Matches   30;  Conservative    6;  Mismatches   17;  Indels   62;  Gaps    6;

Qy      27 QAAILPGKLEICSDDQGNLSPS----------------RETSVDDGLGSQ---------- 60
           ||  |||:| : |   | ||||                |  : | || :|
Db     123 QAGFLPGELGLSS--PGPLSPSGVKSSSQYYTSFPSNPRRRAADGGLDTQPKKVRKVPPG 180

Qy      61 -----YPLKKDSSGDQF--------------LSVPSPF------------TWSKP 84
                ||    |||| :               : ||||             || |
Db     181 LPSSVYP---SSSGDNYSRDATAYPSAKTPSSAYPSPFYVADGSLHPSAELWSPP 232


RESULT 12
NIP2_HUMAN
ID   NIP2_HUMAN     STANDARD;      PRT;   314 AA.
AC   Q12982;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BCL2/adenovirus E1B 19-kDa protein-interacting protein 2.
GN   BNIP2 OR NIP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95042730; PubMed=7954800;
RA   Boyd J.M., Malstrom S., Subramanian T., Venkatesh L.K., Schaeper U.,
RA   Elangovan B., D'Sa-Eipper C., Chinnadurai G.;
RT   "Adenovirus E1B 19 kDa and Bcl-2 proteins interact with a common set
RT   of cellular proteins.";
RL   Cell 79:341-351(1994).
CC   -!- FUNCTION: IMPLICATED IN THE SUPPRESSION OF CELL DEATH. INTERACTS
CC       WITH THE BCL-2 AND ADENOVIRUS E1B 19 kDa PROTEINS.
CC   -!- SUBCELLULAR LOCATION: LOCALIZES TO THE NUCLEAR ENVELOPE REGION AND
CC       TO OTHER CYTOPLASMIC STRUCTURES.
CC   -!- SIMILARITY: CONTAINS 1 CRAL-TRIO DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U15173; AAC00021.1; -.
DR   Genew; HGNC:1083; BNIP2.
DR   MIM; 603292; -.
DR   InterPro; IPR001251; CRAL_TRIO.
DR   SMART; SM00516; SEC14; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
KW   Apoptosis.
FT   DOMAIN      147    304       CRAL-TRIO.
SQ   SEQUENCE   314 AA;  36018 MW;  0A3CEC636569DD2F CRC64;

  Query Match             14.7%;  Score 67.5;  DB 1;  Length 314;
  Best Local Similarity   31.3%;  Pred. No. 5;
  Matches   31;  Conservative    7;  Mismatches   36;  Indels   25;  Gaps    7;

Qy       3 PAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGN------LSPSRETSVDDG 56
           | | |    :| ||   |    |||    || ||:     ||      ::|    ::|
Db      16 PIPLPEDDSIEADILAIT--GPEDQ----PGSLEV----NGNKVRKKLMAPDISLTLDPS 65

Qy      57 LGSQYPLKKDSSGD---QFLSVPS----PFTWSK--PKP 86
             ||      | ||:    |  ||     | |    |||
Db      66 DGSVLSDDLDESGEIDLDGLDTPSENSNEFEWEDDLPKP 104


RESULT 13
VNNL_DROME
ID   VNNL_DROME     STANDARD;      PRT;   558 AA.
AC   Q9NFP1; Q9W431;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Vanin-like protein precursor.
GN   CG3648.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99432008; PubMed=10501839;
RA   Granjeaud S., Naquet P., Galland F.;
RT   "An EST description of the new Vanin gene family conserved from fly to
RT   human.";
RL   Immunogenetics 49:964-972(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

```
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: PROBABLE HYDROLASE.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- TISSUE SPECIFICITY: Expressed in larvae and early pupae.
CC   -!- SIMILARITY: BELONGS TO THE CN HYDROLASE FAMILY. BTD/VNN SUBFAMILY.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ276261; CAB77020.1; -.
DR   EMBL; AE003436; AAF46129.1; ALT_SEQ.
DR   FlyBase; FBgn0040069; vanin-like.
DR   InterPro; IPR003010; Ntlse/CNhydtse.
DR   Pfam; PF00795; CN_hydrolase; 1.
KW   Hydrolase; Signal; Glycoprotein; GPI-anchor.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    532       VANIN-LIKE PROTEIN.
FT   PROPEP      533    558       REMOVED IN MATURE FORM (POTENTIAL).
FT   LIPID       532    532       GPI-ANCHOR (POTENTIAL).
FT   CARBOHYD     65     65       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    354    354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    379    379       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   558 AA;  62341 MW;  7920621923BB7779 CRC64;

  Query Match             14.6%;  Score 67;  DB 1;  Length 558;
  Best Local Similarity   32.8%;  Pred. No. 11;
  Matches   22;  Conservative    5;  Mismatches   22;  Indels   18;  Gaps    3;

Qy      38 CSDD----------QGNLSPSRETSVDDGLGSQYP------LKKDSSGDQFLSV-PSPF 79
           || |          :| |  || |     :|  ||      |  |:  |  | : || |
Db     412 CSGDSIDDCGKLLPTEGELQQSRVTFTRLAIGVTYPESREFLLFPDTLQDSLLPLEPSQF 471

Qy      80 TWSKPKP 86
           ||: ||
Db     472 EWSQRKP 478

RESULT 14
RIK3_MOUSE
ID   RIK3_MOUSE     STANDARD;      PRT;   486 AA.
AC   Q9QZL0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Receptor-interacting serine/threonine protein kinase 3 (EC 2.7.1.-)
DE   (RIP-like protein kinase 3) (Receptor-interacting protein 3) (RIP-3)
DE   (mRIP3).
GN   RIPK3 OR RIP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF ASP-143.
RC   TISSUE=Embryo;
RX   MEDLINE=99421935; PubMed=10490590;
RA   Pazdernik N.J., Donner D.B., Goebl M.G., Harrington M.A.;
RT   "Mouse receptor interacting protein 3 does not contain a caspase-
RT   recruiting or a death domain but induces apoptosis and activates NF-
RT   kappaB.";
RL   Mol. Cell. Biol. 19:6500-6508(1999).
CC   -!- FUNCTION: Promotes apoptosis.
CC   -!- SUBUNIT: Binds TRAF2 and RIPK1 and is recruited to the TNFR-1
CC       signaling complex (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in embryo and in adult spleen,
CC       liver, testis, heart, brain and lung.
CC   -!- PTM: Autophosphorylated.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF178953; AAF03133.1; -.
DR   MGD; MGI:2154952; Ripk3.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Apoptosis.
FT   DOMAIN       22    292       PROTEIN KINASE.
FT   NP_BIND      28     36       ATP (BY SIMILARITY).
FT   BINDING      51     51       ATP (BY SIMILARITY).
FT   ACT_SITE    143    143       PROBABLE.
FT   DOMAIN      352    438       PRO-RICH.
FT   MUTAGEN     143    143       D->N: NO AUTOPHOSPHORYLATION.
SQ   SEQUENCE   486 AA;  53336 MW;  DD264E69187D3436 CRC64;

  Query Match             14.5%;  Score 66.5;  DB 1;  Length 486;
  Best Local Similarity   24.8%;  Pred. No. 11;
  Matches   28;  Conservative   12;  Mismatches   36;  Indels   37;  Gaps    6;

Qy       7 PFQYCVETDIPLKT-------RKTSEDQAAILPGKLEICSDDQ------GNLSPSRETSV 53
           | |   | | | :|       |   |: :  :|||   | : |      |  :|:| ||
Db     325 PSQRGTEMDCPRETMVSKMLDRLHLEEPSGPVPGK---CPERQAQDTSVGPATPAR-TSS 380

Qy      54 DDGLGS-----QYPLKKDSSGDQFLSVPSP--------------FTWSKPKP 86
           |   |:       | :  : |  |   | |              : |: | |
```

```
Db    381 DPVAGTPQIPHTLPFRGTTPGPVFTETPGPHPQRNQGDGRHGTPWYPWTPPNP 433


RESULT 15
HGF_MOUSE
ID   HGF_MOUSE      STANDARD;      PRT;   728 AA.
AC   Q08048; Q64007; Q61662;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hepatocyte growth factor precursor (Scatter factor) (SF)
DE   (Hepatopoeitin A).
GN   HGF.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND SEQUENCE OF 496-504.
RC   TISSUE=Mammary fibroblast;
RX   MEDLINE=94183257; PubMed=8135822;
RA   Sasaki M., Nishio M., Sasaki T., Enami J.;
RT   "Identification of mouse mammary fibroblast-derived mammary growth
RT   factor as hepatocyte growth factor.";
RL   Biochem. Biophys. Res. Commun. 199:772-779(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94363381; PubMed=8081873;
RA   Lee C.C., Kozak C.A., Yamada K.M.;
RT   "Structure, genetic mapping, and expression of the mouse Hgf/scatter
RT   factor gene.";
RL   Cell Adhes. Commun. 1:101-111(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94060105; PubMed=8241272;
RA   Liu Y., Michalopoulos G.K., Zarnegar R.;
RT   "Molecular cloning and characterization of cDNA encoding mouse
RT   hepatocyte growth factor.";
RL   Biochim. Biophys. Acta 1216:299-303(1993).
CC   -!- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL
CC       HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPHIC FACTOR, AND ACTS
CC       AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.
CC       IT HAS NO DETECTABLE PROTEASE ACTIVITY.
CC   -!- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A
CC       DISULFIDE BOND.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMINOGEN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10212; BAA01064.1; -.
DR   EMBL; D10213; BAA01065.1; -.
DR   EMBL; S71816; AAB31855.1; -.
DR   EMBL; X72307; CAA51054.1; ALT_INIT.
DR   HSSP; P14210; 1BHT.
DR   MEROPS; S01.982; -.
DR   MGD; MGI:96079; Hgf.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR003014; PAN.
DR   InterPro; IPR003609; Pan_app.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00024; PAN; 1.
DR   Pfam; PF00051; kringle; 4.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00018; KRINGLE.
DR   ProDom; PD000395; Kringle; 4.
DR   SMART; SM00130; KR; 4.
DR   SMART; SM00473; PAN_AP; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00021; KRINGLE_1; 4.
DR   PROSITE; PS50070; KRINGLE_2; 4.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
KW   Growth factor; Kringle; Glycoprotein; Serine protease homolog;
KW   Repeat; Signal; Alternative splicing.
FT   SIGNAL       1     32       BY SIMILARITY.
FT   CHAIN       33    495       HEPATOCYTE GROWTH FACTOR ALPHA CHAIN.
FT   CHAIN      496    728       HEPATOCYTE GROWTH FACTOR BETA CHAIN.
FT   MOD_RES     33     33       PYRROLIDONE CARBOXYLIC ACID
FT                               (BY SIMILARITY).
FT   DOMAIN      33    128       PAP.
FT   DOMAIN     129    207       KRINGLE 1.
FT   DOMAIN     212    289       KRINGLE 2.
FT   DOMAIN     306    384       KRINGLE 3.
FT   DOMAIN     392    470       KRINGLE 4.
FT   DOMAIN     496    728       SERINE PROTEASE-LIKE.
FT   DISULFID    71     97       BY SIMILARITY.
FT   DISULFID    75     85       BY SIMILARITY.
FT   DISULFID   488    607       INTERCHAIN (BY SIMILARITY).
FT   CARBOHYD   295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   569    569       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   163    167       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   344    344       N -> K (IN REF. 2).
FT   CONFLICT   479    479       V -> L (IN REF. 2).
FT   CONFLICT   564    564       R -> H (IN REF. 3).
SQ   SEQUENCE   728 AA;  82944 MW;  A0381FC497534328 CRC64;

  Query Match             14.4%;  Score 66;  DB 1;  Length 728;
  Best Local Similarity   23.5%;  Pred. No. 20;
  Matches   19;  Conservative   12;  Mismatches   20;  Indels   30;  Gaps    3;

Qy       7 PFQYCV----------ETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDG 56
           |::||           |||:|::| :               |   ||        :: :|
Db     280 PWEYCAIKTCAHSAVNETDVPMETTE--------------CIQGQGEGYRGTSNTIWNG 324

Qy      57 L-----GSQYPLKKDSSGDQF 72
           :     |||| | | : : |
Db     325 IPCQRWDSQYPHKHDITPENF 345


Search completed: April 11, 2003, 14:22:42
Job time : 16 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:20:32 ; Search time 29 Seconds
(without alignments)
611.036 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	194	42.3	884	13	Q8QHM1	Q8qhm1 xenopus lae
2	129	28.1	226	11	Q9CRU3	Q9cru3 mus musculu
3	126	27.5	901	11	Q9EP91	Q9ep91 mus musculu
4	122	26.6	845	4	Q96H68	Q96h68 homo sapien
5	96.5	21.0	703	11	Q9DBQ6	Q9dbq6 mus musculu
6	96.5	21.0	717	11	Q9JHD3	Q9jhd3 mus musculu
7	84.5	18.4	744	11	O35673	O35673 mus musculu
8	81.5	17.8	329	10	O65389	O65389 arabidopsis
9	79	17.2	1065	10	Q9FJL6	Q9fjl6 arabidopsis
10	77	16.8	1257	5	Q9I7N6	Q9i7n6 drosophila
11	75.5	16.4	595	4	Q9HCI7	Q9hci7 homo sapien
12	73.5	16.0	3484	5	P91257	P91257 caenorhabdi
13	73	15.9	667	2	Q54232	Q54232 streptococc
14	72	15.7	2130	16	Q8Z4P9	Q8z4p9 salmonella
15	71.5	15.6	792	11	Q9R1J1	Q9r1j1 mus musculu
16	71.5	15.6	792	11	Q9Z1J2	Q9z1j2 mus musculu
17	70.5	15.4	348	4	Q9BSL3	Q9bsl3 homo sapien
18	70.5	15.4	401	4	Q9BYS3	Q9bys3 homo sapien
19	70.5	15.4	401	4	Q96NA2	Q96na2 homo sapien
20	70.5	15.4	652	11	Q8VCY4	Q8vcy4 mus musculu
21	70	15.3	204	2	Q9RQL4	Q9rql4 streptococc
22	70	15.3	452	2	Q9S549	Q9s549 streptococc
23	70	15.3	470	2	O07041	O07041 streptococc
24	70	15.3	648	11	Q922S2	Q922s2 mus musculu
25	70	15.3	651	11	Q9CYJ4	Q9cyj4 mus musculu
26	70	15.3	661	2	Q9RIN5	Q9rin5 streptococc
27	68.5	14.9	444	11	Q9D2Q3	Q9d2q3 mus musculu
28	68	14.8	661	2	O52230	O52230 streptococc
29	67	14.6	177	16	Q8XA52	Q8xa52 escherichia
30	67	14.6	1112	11	P97879	P97879 rattus norv
31	67	14.6	1499	13	Q9DDA6	Q9dda6 xenopus lae
32	66.5	14.5	1022	5	O45813	O45813 caenorhabdi
33	66	14.4	363	5	Q9GV96	Q9gv96 toxoplasma
34	66	14.4	430	4	Q15786	Q15786 homo sapien
35	66	14.4	500	15	Q8UTN3	Q8utn3 human immun
36	66	14.4	1286	4	Q9P2C3	Q9p2c3 homo sapien
37	65.5	14.3	246	15	Q9IM52	Q9im52 human immun
38	65.5	14.3	494	15	Q8UPQ3	Q8upq3 human immun
39	65.5	14.3	3105	12	Q8UZI9	Q8uzi9 cercopithic
40	65	14.2	119	15	Q72900	Q72900 human immun
41	65	14.2	175	11	Q922A4	Q922a4 mus musculu
42	65	14.2	280	16	Q9L1M1	Q9l1m1 streptomyce
43	65	14.2	329	11	Q9D0P7	Q9d0p7 mus musculu
44	65	14.2	329	11	Q9QXY5	Q9qxy5 mus musculu
45	65	14.2	414	4	Q96KB7	Q96kb7 homo sapien

ALIGNMENTS

RESULT 1
Q8QHM1
ID Q8QHM1 PRELIMINARY; PRT; 884 AA.
AC Q8QHM1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Nuclear factor of activated T-cell.
GN NF-AT.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Saneyoshi T., Kume S., Mikoshiba K.;
RT "Xenopus laevis, Nuclear Factor of Activated T-cell.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB037471; BAB85496.1; -.
SQ SEQUENCE 884 AA; 97693 MW; F46E49B7E54FD84C CRC64;

Query Match 42.3%; Score 194; DB 13; Length 884;
Best Local Similarity 52.3%; Pred. No. 1.8e-13;
Matches 46; Conservative 10; Mismatches 28; Indels 4; Gaps 1;

```
Qy      2 SPAPFP----FQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGL 57
          || | |    :    || |||||||||| :| | || |: ::|   |||  :   |:
Db    294 SPHPAPSLSLYPCWQETAIPLKTRKTSEDHSARLAGKTELNTEDLACLSPIVDMPADETS 353

Qy     58 GSQYPLKKDSSGDQFLSVPSPFTWSKPK 85
           ||: | :||| |||||||| |||:| |
Db    354 CSQHLLVQDSSTDQFLSVPSHFTWNKHK 381
```

RESULT 2
Q9CRU3
ID Q9CRU3 PRELIMINARY; PRT; 226 AA.
AC Q9CRU3;

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   3110041H08Rik protein (Fragment).
GN   NFATC4 OR 3110041H08RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK014164; BAB29185.1; -.
DR   MGD; MGI:1920431; Nfatc4.
FT   NON_TER       1       1
SQ   SEQUENCE   226 AA;  23584 MW;  882F2BBD4FEDB595 CRC64;

  Query Match             28.1%;  Score 129;  DB 11;  Length 226;
  Best Local Similarity   34.1%;  Pred. No. 9e-07;
  Matches   29;  Conservative   18;  Mismatches   32;  Indels    6;  Gaps    2;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :     |: || |||:|| :||  ||      |::  : :    :  :| : :
Db     138 SPGPFDYVGAPPTESIPQKTRRTSSEQAVALPR-----SEEPPSCNGKLPSGTEDSVAAP 192

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             |:|: :|  :|:||||  ||| :
Db     193 GALRKEVAGMDYLAVPSPLAWSKAR 217


RESULT 3
Q9EP91
ID   Q9EP91      PRELIMINARY;      PRT;   901 AA.
AC   Q9EP91;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcription complex subunit NF-ATc4.
GN   NFATC4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV/EV;
RA   Graef I.A., Crabtree G.R.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C129SVEV;
RA   Graef I.A., Crabtree G.R.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF309389; AAG39446.1; -.
DR   EMBL; AF309388; AAG39446.1; JOINED.
DR   EMBL; AF283284; AAF98174.1; -.
DR   MGD; MGI:1920431; Nfatc4.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01833; TIG; 1.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
DR   PROSITE; PS50254; REL_2; 1.
SQ   SEQUENCE   901 AA;  95820 MW;  7F66A93E76FDB773 CRC64;

  Query Match             27.5%;  Score 126;  DB 11;  Length 901;
  Best Local Similarity   32.9%;  Pred. No. 1.1e-05;
  Matches   28;  Conservative   19;  Mismatches   32;  Indels    6;  Gaps    2;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :     |: || |||:|| :||  ||      |::  : :    :  :| : :
Db     311 SPGPFDYVGAPPTESIPQKTRRTSSEQAVALPR-----SEEPPSCNGKLPSGTEDSVAAP 365

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             |:|: :|  :|::|||  ||| :
Db     366 GALRKEVAGMDYLTMPSPLAWSKAR 390


RESULT 4
Q96H68
ID   Q96H68      PRELIMINARY;      PRT;   845 AA.
AC   Q96H68;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Unknown (Protein for IMAGE:4109469) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUSCLE;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC008857; AAH08857.1; -.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF01833; TIG; 1.
DR   PROSITE; PS50254; REL_2; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   845 AA;  89682 MW;  257F079E2DEA5F16 CRC64;

  Query Match             26.6%;  Score 122;  DB 4;  Length 845;
  Best Local Similarity   40.0%;  Pred. No. 2.8e-05;
  Matches   34;  Conservative   11;  Mismatches   34;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :      : || |||:|| :||  ||   |  | : | |    | ||    ||
Db     254 SPGPFDYVGAPPAESIPQKTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 311

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             :|: :|  :|:||||  ||| :
Db     312 ---RKEVAGMDYLAVPSPLAWSKAR 333


RESULT 5
Q9DBQ6
ID   Q9DBQ6      PRELIMINARY;      PRT;   703 AA.
AC   Q9DBQ6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Nuclear factor of activated T-cells, cytoplasmic 1.
GN   NFATC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LUNG;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004810; BAB23581.1; -.
DR   MGD; MGI:102469; Nfatc1.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS50254; REL_2; 1.
SQ   SEQUENCE   703 AA;  76691 MW;  D737C9B86E280C7D CRC64;

  Query Match             21.0%;  Score 96.5;  DB 11;  Length 703;
  Best Local Similarity   30.4%;  Pred. No. 0.018;
  Matches   24;  Conservative   17;  Mismatches   19;  Indels   19;  Gaps    3;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYP--------LKKDS 67
           :|:|:|||: : |  :  |:|   :| |   |:          | :|         |:| :
Db     319 VPIKSRKTALEHAPSVALKVEPAGEDLGTTPPT----------SDFPPEEYTFQHLRKGA 368

Qy      68 SGDQFLSVP-SPFTWSKPK 85
             :|:|||| : : |:|||
Db     369 FCEQYLSVPQASYQWAKPK 387


RESULT 6
Q9JHD3
ID   Q9JHD3      PRELIMINARY;      PRT;   717 AA.
AC   Q9JHD3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Nuclear factor of activated T cells c.
GN   NFATC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98049829; PubMed=9388475;
RA   Pan S., Koyano-Nakagawa N., Tsuruta L., Amasaki Y., Yokota T.,
RA   Mori S., Arai N., Arai K.;
RT   "Molecular cloning and functional characterization of murine cDNA
RT   encoding transcription factor NFATc.";
RL   Biochem. Biophys. Res. Commun. 240:314-323(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pan S.;.
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF239169; AAF40225.1; -.
DR   MGD; MGI:102469; Nfatc1.
DR   InterPro; IPR002909; IPT_TIG.
DR   InterPro; IPR000451; NF_Rel_dor_fam.
DR   Pfam; PF00554; RHD; 1.
DR   Pfam; PF01833; TIG; 1.
DR   SMART; SM00429; IPT; 1.
DR   PROSITE; PS50254; REL_2; 1.
SQ   SEQUENCE   717 AA;  77936 MW;  1D18F1EF5A5135C6 CRC64;

  Query Match             21.0%;  Score 96.5;  DB 11;  Length 717;
  Best Local Similarity   30.4%;  Pred. No. 0.018;
  Matches   24;  Conservative   17;  Mismatches   19;  Indels   19;  Gaps    3;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYP--------LKKDS 67
           :|:|:|||: : |  :  |:|   :| |   |:          | :|         |:| :
Db     333 VPIKSRKTALEHAPSVALKVEPAGEDLGTTPPT----------SDFPPEEYTFQHLRKGA 382

Qy      68 SGDQFLSVP-SPFTWSKPK 85
             :|:|||| : : |:|||
Db     383 FCEQYLSVPQASYQWAKPK 401


RESULT 7
O35673
ID   O35673      PRELIMINARY;      PRT;   744 AA.
AC   O35673;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   MSTK2S kinase-like protein.
GN   NEK4 OR MSTK2S.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=20001940; PubMed=10529384;
RA   Hayashi K., Igarashi H., Ogawa M., Sakaguchi N.;
RT   "Activity and substrate specificity of the murine STK2
RT   serine/threonine kinase that is structurally related to the mitotic
RT   regulator protein NIMA of Aspergillus nidulans.";
RL   Biochem. Biophys. Res. Commun. 264:449-456(1999).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; Y09234; CAA70436.1; -.
DR   HSSP; Q63450; 1A06.
DR   MGD; MGI:1344404; Nek4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   744 AA;  83542 MW;  82E1AAF3E9DCC72A CRC64;

  Query Match             18.4%;  Score 84.5;  DB 11;  Length 744;
  Best Local Similarity   41.2%;  Pred. No. 0.45;
  Matches   21;  Conservative    3;  Mismatches   22;  Indels    5;  Gaps    1;

Qy      39 SDDQGNLSPSRETSVDDGLGSQ-----YPLKKDSSGDQFLSVPSPFTWSKP 84
           ||  ||| | | :   || ||:     ||  ||   |||  :|     | |
Db     418 SDQPGNLLPRRSSDGGDGEGSELVKPLYPSNKDQKPDQFQELPPRLLPSYP 468
```

```
RESULT 8
O65389
ID   O65389      PRELIMINARY;      PRT;   329 AA.
AC   O65389;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F12F1.23 protein (Hypothetical 36.7 kDa protein).
GN   F12F1.23.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Vysotskaia V.S., Osborne B.I., Schwartz J.R., Toriumi M., Yu G.,
RA   Kwan A., Oji O., Liu S., Buehler E., Conway A.B., Conway A.R.,
RA   Dewar K., Feng J., Kim C., Kurtz D., Palm C.J., Li Y., Shinn P.,
RA   Sun H., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA   Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA   Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA   Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Full Length cDNA of gene F12F1.23 (GI:3157936).";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA   Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA   Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA   Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA   Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene F12F1.23 (GI:3157936).";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002131; AAC17619.1; -.
DR   EMBL; AF360349; AAK28646.1; -.
DR   EMBL; AY051085; AAK93762.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   329 AA;  36742 MW;  4A87A14ECA5A41CC CRC64;

  Query Match             17.8%;  Score 81.5;  DB 10;  Length 329;
  Best Local Similarity   25.3%;  Pred. No. 0.37;
  Matches   20;  Conservative   10;  Mismatches   26;  Indels   23;  Gaps    1;

Qy      27 QAAILPGKLEICSDDQGNLSPSRETSVDDGLGS----------------------QYPL 63
           |::    :  :|    ||:  :   |:||| ||                      :||:
Db      20 QSSTTSSQSNLCRSSCGNIPINYPFSIDDGCGSPYYRHMLICSDNDTKLELRTPSGKYPV 79

Qy      64 KKDSSGDQFLSVPSPFTWS 82
           |  |  |  | |  || |:
Db      80 KSISYSDPHLLVSDPFMWN 98


RESULT 9
Q9FJL6
ID   Q9FJL6      PRELIMINARY;      PRT;  1065 AA.
AC   Q9FJL6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Gb|AAD25584.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=98403884; PubMed=9734815;
RA   Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RT   Sequence features of the regions of 1,367,185 bp covered by 19
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:203-216(1998).
DR   EMBL; AB013396; BAB08870.1; -.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR004274; NIF.
DR   Pfam; PF00533; BRCT; 1.
DR   Pfam; PF03031; NIF; 1.
DR   SMART; SM00292; BRCT; 1.
DR   PROSITE; PS50172; BRCT; 1.
SQ   SEQUENCE   1065 AA;  120932 MW;  689B6FDFBF486CF0 CRC64;

  Query Match             17.2%;  Score 79;  DB 10;  Length 1065;
  Best Local Similarity   33.9%;  Pred. No. 2.9;
  Matches   20;  Conservative   11;  Mismatches   26;  Indels    2;  Gaps    1;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGS 59
             |||    |:  | :| |: :  :|:| || |  :|:  ||   :  || |   :| : |
Db     617 LSPPTNKFKMSVASDSPVHSSSSSDDLAAFLDAELDSASD--ASSGPSEEEEAEDDVES 673


RESULT 10
Q9I7N6
ID   Q9I7N6      PRELIMINARY;      PRT;  1257 AA.
AC   Q9I7N6; Q9TVE8; Q9TVE9; Q9Y195; Q9UA38; Q9U971;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   DNA-repair protein complementing XP-G cells homolog (Xeroderma
DE   pigmentosum group G complementing protein homolog) (Mutagen-sensitive
DE   201) (Nucleotide excision repair protein mus201) (DNA repair
DE   endonuclease).
GN   MUS201 OR XPG OR CG10890.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RC   TISSUE=EMBRYO;
RX   MEDLINE=99326510; PubMed=10395909;
RA   Houle J.-F., Friedberg E.C.;
RT   "The Drosophila ortholog of the human XPG gene.";
RL   Gene 234:353-360(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RC   TISSUE=EMBRYO;
RX   MEDLINE=20274058; PubMed=10812334;
RA   Sekelsky J.J., Hollis K.J., Eimerl A.I., Burtis K.C., Hawley R.S.;
RT   "Nucleotide excision repair endonuclease genes in Drosophila
RT   melanogaster.";
RL   Mutat. Res. 459:219-228(2000).
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RA   Calleja F.M.G.R., Nivard M.J.M., Vogel E.W., Eeken J.C.J.;
RT   "Genetic effects in the Drosophila mutant mus201[D1] are caused by
RT   alterations in the Drosophila homolog of the nucleotide excision
RT   repair gene encoding XPG.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [4]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [5]
RP   SEQUENCE OF 880-1257 FROM N.A.
RC   TISSUE=ANTENNA;
RA   Hovemann B.T., Heiermann R., Malz J., Richard A., Stoertkuhl K.F.,
RA   Sehlmeyer F.;
RT   "Deliniation of a modular structured enhancer that gives rise to
RT   olfactory organ specific expression of Drosophilal-acylglycerol-3-
RT   phosphate O-acyltransferase.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE INVOLVED IN DNA
CC       EXCISION REPAIR. MAKES THE 3'INCISION IN DNA NUCLEOTIDE EXCISION
CC       REPAIR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LONG (SHOWN HERE) AND SHORT; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: TO OTHER MEMBERS OF THE XPG/RAD2 FAMILY. BELONGS TO
CC       THE RAD2(S.CEREVISIAE)/XPG SUBFAMILY.
DR   EMBL; AF125578; AAD47568.1; -.
DR   EMBL; AF144092; AAD38375.1; -.
DR   EMBL; AF162792; AAD50779.1; -.
DR   EMBL; AF162793; AAD50780.1; -.
DR   EMBL; AE003621; AAF52678.1; -.
DR   EMBL; AE003621; AAG22424.1; -.
DR   EMBL; AJ389385; CAB52135.1; -.
DR   FlyBase; FBgn0002887; mus201.
DR   InterPro; IPR000513; Exo_N_I.
DR   InterPro; IPR003584; HHH_2.
DR   InterPro; IPR002950; Josephin.
DR   InterPro; IPR003903; UIM.
DR   InterPro; IPR001532; XPGC_Rad.
DR   Pfam; PF02809; UIM; 1.
DR   Pfam; PF00867; XPG_I; 1.
DR   Pfam; PF00752; XPG_N; 1.
DR   PRINTS; PR00853; XPGRADSUPER.
DR   SMART; SM00279; HhH2; 1.
DR   SMART; SM00484; XPGI; 1.
DR   SMART; SM00485; XPGN; 1.
DR   PROSITE; PS00841; XPG_1; FALSE_NEG.
DR   PROSITE; PS00842; XPG_2; 1.
KW   DNA repair; DNA-binding; Nuclear protein; Hydrolase; Nuclease;
KW   Endonuclease; Alternative splicing.
FT   DOMAIN        1    116       N-DOMAIN.
FT   DOMAIN      885    994       I-DOMAIN.
FT   VARSPLIC     30     50       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    183    183       D -> N (IN REF. 3 AND 5).
FT   CONFLICT    449    449       A -> T (IN REF. 3 AND 5).
FT   CONFLICT    501    501       K -> N (IN REF. 1).
FT   CONFLICT    573    573       D -> N (IN REF. 1).
FT   CONFLICT    580    580       D -> G (IN REF. 1).
FT   CONFLICT    634    634       E -> D (IN REF. 1).
FT   CONFLICT    693    694       SV -> NA (IN REF. 1).
FT   CONFLICT    705    705       T -> A (IN REF. 1).
FT   CONFLICT    831    831       I -> L (IN REF. 1).
FT   CONFLICT    846    846       E -> G (IN REF. 1).
FT   CONFLICT    851    851       L -> M (IN REF. 1).
FT   CONFLICT    996    996       Q -> R (IN REF. 3).
SQ   SEQUENCE   1257 AA;  141424 MW;  4F2A77B46457F68B CRC64;

  Query Match             16.8%;  Score 77;  DB 5;  Length 1257;
  Best Local Similarity   25.6%;  Pred. No. 6.1;
  Matches   20;  Conservative   14;  Mismatches   30;  Indels   14;  Gaps    2;

Qy     23 TSEDQAAILPGK-------LEICSDDQG-------NLSPSRETSVDDGLGSQYPLKKDSS 68
          :|:|:: :   |        :|:|  |          | :||:  |: |   : | :|:
Db    699 SSDDESTLKSSKIVPKQELIELCDSDDNKNNRLSPNKTPSKNKSIKDFFETSYVVKRTPD 758

Qy     69 GDQFLSVPSPFTWSKPKP 86
          |  :  || |   |||
Db    759 KSQASNETSPGTPKTPKP 776


RESULT 11
Q9HCI7
ID   Q9HCI7      PRELIMINARY;      PRT;   595 AA.
AC   Q9HCI7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1585 protein (Fragment).
GN   KIAA1585.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20450683; PubMed=10997877;
RA   Nagase T., Kikuno R., Nakayama M., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.
RT   XVIII. The complete sequences of 100 new cDNA clones from brain which
RT   code for large proteins in vitro.";
RL   DNA Res. 7:273-281(2000).
DR   EMBL; AB046805; BAB13411.1; -.
DR   InterPro; IPR001841; Znf_ring.
FT   NON_TER       1       1
SQ   SEQUENCE   595 AA;  64197 MW;  3BDBECC083B606D2 CRC64;

  Query Match             16.4%;  Score 75.5;  DB 4;  Length 595;
  Best Local Similarity   31.1%;  Pred. No. 3.7;
  Matches   23;  Conservative   14;  Mismatches   32;  Indels    5;  Gaps    2;

Qy     10 YCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ----YPLKK 65
          :| ||: |  :  |     : ||   |  :| | :|||  |:::   :||      |
Db    157 FCEETEKPSDSSFTLCLTHSPLPSTSEPTTDPQASLSPMSESTLSIAIGSSVINGLPTYN 216
```

```
Qy      66 DSSGDQF-LSVPSP 78
           | |:| :::|||
Db     217 GLSIDRFGINIPSP 230


RESULT 12
P91257
ID   P91257      PRELIMINARY;      PRT;  3484 AA.
AC   P91257;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 385.7 kDa protein.
GN   F12F3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80022; AAC25885.2; -.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00041; fn3; 7.
DR   Pfam; PF00047; ig; 10.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 8.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 8.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Hypothetical protein; Immunoglobulin domain; Repeat.
SQ   SEQUENCE   3484 AA;  385746 MW;  21DD6F6893020F4F CRC64;

  Query Match             16.0%;  Score 73.5;  DB 5;  Length 3484;
  Best Local Similarity   34.4%;  Pred. No. 52;
  Matches   22;  Conservative    9;  Mismatches   30;  Indels    3;  Gaps    2;

Qy      15 DIPLKTRKTSEDQAAI-LPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFL 73
           |:  |::||:|||  :    |||  :|    :    || |||    :   ||    | |:
Db     693 DVTDKSKKTTEDQTKVATDSKLEKAADTTKQI--ETETVVDDKSKKKVLKKKTEKSDSFI 750

Qy      74 SVPS 77
           |  |
Db     751 SQKS 754


RESULT 13
Q54232
ID   Q54232      PRELIMINARY;      PRT;  667 AA.
AC   Q54232;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Oligopeptide-binding lipoprotein precursor.
GN   HPPA.
OS   Streptococcus gordonii.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Streptococcaceae; Streptococcus.
OX   NCBI_TaxID=1302;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DL1 CHALLIS;
RX   MEDLINE=96125230; PubMed=8550445;
RA   Jenkinson H.F., Baker R.A., Tannock G.W.;
RT   "A binding-lipoprotein-dependent oligopeptide transport system in
RT   Streptococcus gordonii essential for uptake of hexa- and
RT   heptapeptides.";
RL   J. Bacteriol. 178:68-77(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DL1 CHALLIS;
RA   Jenkinson H.F.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L41358; AAB46614.1; -.
DR   InterPro; IPR000914; SBP_bac_5.
DR   Pfam; PF00496; SBP_bac_5; 1.
DR   PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW   Lipoprotein; Signal.
FT   SIGNAL        1     22       POTENTIAL.
SQ   SEQUENCE   667 AA;  74003 MW;  5CF994ECB920BAB6 CRC64;

  Query Match             15.9%;  Score 73;  DB 2;  Length 667;
  Best Local Similarity   31.7%;  Pred. No. 8.1;
  Matches   26;  Conservative   11;  Mismatches   29;  Indels   16;  Gaps    4;

Qy       8 FQYCVETD------IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRE---TSVDDGLG 58
           | |  | |      :      |::  | :: | ||  :|: ||| ||     |   |||
Db      42 FNYIYEVDPENLNYLISSKAATTDLTANLIDGLLE--NDNYGNLVPSMAEDWTVSKDGLT 99

Qy      59 SQYPLKKD-----SSGDQFLSV 75
             | |:||     | |:::  |
Db     100 YTYTLRKDAKWYTSDGEEYADV 121


RESULT 14
Q8Z4P9
ID   Q8Z4P9      PRELIMINARY;      PRT;  2130 AA.
AC   Q8Z4P9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative exported protein.
GN   STY2760.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
```

```
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627275; CAD02718.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   2130 AA;  228149 MW;  DFA535A11DF545D9 CRC64;

  Query Match             15.7%;  Score 72;  DB 16;  Length 2130;
  Best Local Similarity   27.0%;  Pred. No. 43;
  Matches   24;  Conservative   13;  Mismatches   38;  Indels   14;  Gaps    2;

Qy       5 PFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSV---------- 53
           | |::|      | | ||  | | |    :: : |:|    ||   | |
Db     501 PLPYRYMELFIDPAKNRK-GEHQDAWDNQRVTVSSEDMRASSPEHYTGVTDVNGQAHLTL 559

Qy      54 --DDGLGSQYPLKKDSSGDQFLSVPSPFT 80
             | |:| : |::     |:  :|  ||:
Db     560 QHDSGMGGETPIRIVMPDDEGGNVELPFS 588


RESULT 15
Q9R1J1
ID   Q9R1J1      PRELIMINARY;      PRT;   792 AA.
AC   Q9R1J1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Serine/threonine-protein kinase NEK4.
GN   NEK4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99321807; PubMed=10393247;
RA   Chen A., Yanai A., Arama E., Kilfin G., Motro B.;
RT   "NIMA-related kinases: isolation and characterization of murine nek3
RT   and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3.";
RL   Gene 234:127-137(1999).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AF099067; AAD16287.1; -.
DR   HSSP; Q63450; 1A06.
DR   MGD; MGI:1344404; Nek4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   792 AA;  89093 MW;  9B8F373117377CC6 CRC64;

  Query Match             15.6%;  Score 71.5;  DB 11;  Length 792;
  Best Local Similarity   44.7%;  Pred. No. 15;
  Matches   17;  Conservative    2;  Mismatches   14;  Indels    5;  Gaps    1;

Qy      39 SDDQGNLSPSRETSVDDGLGSQ-----YPLKKDSSGDQ 71
           ||  ||| | | :   || ||:     ||  ||   ||
Db     418 SDQPGNLLPRRSSDGGDGEGSELVKPLYPSNKDQKPDQ 455


Search completed: April 11, 2003, 14:23:19
Job time : 33 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:21:47 ; Search time 14 Seconds
(without alignments)
180.741 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	430	93.7	708	1	US-08-396-479B-8	Sequence 8, Appli
2	430	93.7	708	1	US-08-818-823-8	Sequence 8, Appli
3	430	93.7	739	1	US-08-396-479B-10	Sequence 10, Appl
4	430	93.7	739	1	US-08-818-823-10	Sequence 10, Appl
5	430	93.7	1068	1	US-08-396-479B-12	Sequence 12, Appl
6	430	93.7	1068	1	US-08-818-823-12	Sequence 12, Appl
7	430	93.7	1075	5	PCT-US94-07297-41	Sequence 41, Appl
8	122	26.6	902	1	US-08-396-479B-6	Sequence 6, Appli
9	122	26.6	902	1	US-08-818-823-6	Sequence 6, Appli
10	114	24.8	716	1	US-08-396-479B-4	Sequence 4, Appli
11	114	24.8	716	1	US-08-818-823-4	Sequence 4, Appli
12	114	24.8	716	3	US-09-037-190-38	Sequence 38, Appl
13	114	24.8	716	3	US-09-037-190-46	Sequence 46, Appl
14	114	24.8	716	3	US-09-037-192-38	Sequence 38, Appl
15	114	24.8	716	3	US-09-037-192-46	Sequence 46, Appl
16	114	24.8	716	4	US-09-037-143-38	Sequence 38, Appl
17	114	24.8	716	4	US-09-037-143-46	Sequence 46, Appl
18	114	24.8	716	4	US-09-049-691-38	Sequence 38, Appl
19	114	24.8	716	4	US-09-049-691-46	Sequence 46, Appl
20	114	24.8	716	4	US-08-260-174-38	Sequence 38, Appl
21	114	24.8	716	4	US-08-260-174-46	Sequence 46, Appl
22	114	24.8	716	4	US-09-338-128A-38	Sequence 38, Appl
23	114	24.8	716	4	US-09-338-128A-46	Sequence 46, Appl
24	114	24.8	716	4	US-09-232-346-38	Sequence 38, Appl
25	114	24.8	716	4	US-09-232-346-46	Sequence 46, Appl
26	114	24.8	716	4	US-09-037-192-38	Sequence 38, Appl
27	114	24.8	716	4	US-09-037-192-46	Sequence 46, Appl
28	114	24.8	716	5	PCT-US94-07297-37	Sequence 37, Appl
29	114	24.8	761	2	US-08-124-981A-2	Sequence 2, Appli
30	70	15.3	643	2	US-08-245-511-47	Sequence 47, Appl
31	70	15.3	643	2	US-08-600-993A-47	Sequence 47, Appl
32	67.5	14.7	314	2	US-08-408-095-23	Sequence 23, Appl
33	64	13.9	422	4	US-08-790-186A-4	Sequence 4, Appli
34	63	13.7	398	3	US-09-045-632-31	Sequence 31, Appl
35	63	13.7	441	3	US-09-045-632-37	Sequence 37, Appl
36	63	13.7	498	3	US-09-045-632-30	Sequence 30, Appl
37	63	13.7	541	3	US-09-045-632-36	Sequence 36, Appl
38	63	13.7	599	3	US-09-045-632-28	Sequence 28, Appl
39	63	13.7	642	3	US-09-045-632-35	Sequence 35, Appl
40	63	13.7	818	3	US-09-045-632-25	Sequence 25, Appl
41	63	13.7	861	3	US-09-045-632-34	Sequence 34, Appl
42	63	13.7	918	3	US-09-045-632-21	Sequence 21, Appl
43	63	13.7	961	3	US-09-045-632-33	Sequence 33, Appl
44	63	13.7	1018	3	US-09-045-632-16	Sequence 16, Appl
45	63	13.7	1061	3	US-09-045-632-32	Sequence 32, Appl

ALIGNMENTS

```
RESULT 1
US-08-396-479B-8
; Sequence 8, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  708 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-396-479B-8

  Query Match             93.7%;  Score 430;  DB 1;  Length 708;
  Best Local Similarity   91.9%;  Pred. No. 2e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || |||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
```

```
Db    381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 2
US-08-818-823-8
; Sequence 8, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  708 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-818-823-8

  Query Match             93.7%;  Score 430;  DB 1;  Length 708;
  Best Local Similarity   91.9%;  Pred. No. 2e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || |||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
Db     381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 3
US-08-396-479B-10
; Sequence 10, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  739 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-396-479B-10

  Query Match             93.7%;  Score 430;  DB 1;  Length 739;
  Best Local Similarity   91.9%;  Pred. No. 2.1e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || |||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
Db     381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 4
US-08-818-823-10
; Sequence 10, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
```

```
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  739 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-818-823-10

  Query Match             93.7%;  Score 430;  DB 1;  Length 739;
  Best Local Similarity   91.9%;  Pred. No. 2.1e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
          | || ||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db    321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy     61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
          ||||||| ||||||||||||||||||
Db    381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 5
US-08-396-479B-12
; Sequence 12, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-396-479B-12

  Query Match             93.7%;  Score 430;  DB 1;  Length 1068;
  Best Local Similarity   91.9%;  Pred. No. 3.5e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
          | || ||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db    321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy     61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
          ||||||| ||||||||||||||||||
Db    381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 6
US-08-818-823-12
; Sequence 12, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-818-823-12

  Query Match             93.7%;  Score 430;  DB 1;  Length 1068;
  Best Local Similarity   91.9%;  Pred. No. 3.5e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
          | || ||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db    321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy     61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
          ||||||| ||||||||||||||||||
Db    381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 7
PCT-US94-07297-41
; Sequence 41, Application PC/TUS9407297
;  GENERAL INFORMATION:
;    APPLICANT:  Arai, Naoko
;    APPLICANT:  Masuda, Esteban S.
;    APPLICANT:  Tokumitsu, Hiroshi
;    TITLE OF INVENTION:  PURIFIED COMPONENTS OF MAMMALIAN
;    TITLE OF INVENTION:  TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John H. C. Blasdale, Schering-Plough Corporation, M-3-W
;      STREET:  One Giralda Farms
;      CITY:  Madison
;      STATE:  New Jersey
```

```
;       COUNTRY:  USA
;       ZIP:  07940-1000
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Apple Macintosh IIci
;       OPERATING SYSTEM:  System Software 7.1
;       SOFTWARE:  Microsoft Word 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/07297
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/222,626
;       FILING DATE:  04-APR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/148,061
;       FILING DATE:  05-NOV-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/113,971
;       FILING DATE:  30-AUG-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/099,998
;       FILING DATE:  30-JUL-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/088,483
;       FILING DATE:  06-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Blasdale, John H. C.
;       REGISTRATION NUMBER:  31,895
;       REFERENCE/DOCKET NUMBER:  DX0392K4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-822-7398
;       TELEFAX:  201-822-7039
;   INFORMATION FOR SEQ ID NO:  41:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1075 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
PCT-US94-07297-41

  Query Match             93.7%;  Score 430;  DB 5;  Length 1075;
  Best Local Similarity   91.9%;  Pred. No. 3.5e-47;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || |||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
Db     381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 8
US-08-396-479B-6
; Sequence 6, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  902 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-396-479B-6

  Query Match             26.6%;  Score 122;  DB 1;  Length 902;
  Best Local Similarity   40.0%;  Pred. No. 3.5e-07;
  Matches   34;  Conservative   11;  Mismatches   34;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
             || || :       : || |||:|| :||  ||   |  | : | |    | ||    ||
Db     311 SPGPFDYVGAPPAESIPQKTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 368

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
              :|: :|  :|:||||  ||| :
Db     369 ---RKEVAGMDYLAVPSPLAWSKAR 390


RESULT 9
US-08-818-823-6
; Sequence 6, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  902 amino acids
;      TYPE:  amino acid
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-818-823-6

  Query Match             26.6%;  Score 122;  DB 1;  Length 902;
  Best Local Similarity   40.0%;  Pred. No. 3.5e-07;
  Matches   34;  Conservative   11;  Mismatches   34;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :      : || |||:|| :||  ||   |  | : | |    | ||    ||
Db     311 SPGPFDYVGAPPAESIPQKTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 368

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             :|: :|  :|:|||| ||| :
Db     369 ---RKEVAGMDYLAVPSPLAWSKAR 390


RESULT 10
US-08-396-479B-4
; Sequence 4, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  716 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-396-479B-4

  Query Match             24.8%;  Score 114;  DB 1;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 2.8e-06;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 11
US-08-818-823-4
; Sequence 4, Application US/08818823
; Patent No. 5708158
; GENERAL INFORMATION:
;   APPLICANT:  HOEY, Timothy
;   TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;   NUMBER OF SEQUENCES:  18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;     STREET:  4 Embarcadero Center, Suite 3400
;     CITY:  San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94111
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/818,823
;     FILING DATE:  14-MAR-1997
;     CLASSIFICATION:  536
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/396,479
;     FILING DATE:  02-MAR-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Osman, Richard A
;     REGISTRATION NUMBER:  36,627
;     REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 494-8700
;     TELEFAX:  (415) 494-8771
;     TELEX:  210 277299
; INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  716 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-818-823-4

  Query Match             24.8%;  Score 114;  DB 1;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 2.8e-06;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 12
US-09-037-190-38
; Sequence 38, Application US/09037190
; Patent No. 6096515
; GENERAL INFORMATION:
;   APPLICANT:  Crabtree, Gerald R.
;   APPLICANT:  No. 6096515throp, Jeffrey P.
;   APPLICANT:  Ho, Steffan M.
;   TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;   NUMBER OF SEQUENCES:  52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;     STREET:  One Post Office Square
;     CITY:  Boston
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  02109-2170
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/037,190
;       FILING DATE:  09-MAR-1998
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/260,174
;       FILING DATE:  13-JUN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/124,981
;       FILING DATE:  20-SEP-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Vincent, Matthew P.
;       REGISTRATION NUMBER:  36,709
;       REFERENCE/DOCKET NUMBER:  APV-332.03
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-832-1000
;       TELEFAX:  617-832-7000
;   INFORMATION FOR SEQ ID NO:  38:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  716 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-09-037-190-38

  Query Match             24.8%;  Score 114;  DB 3;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 2.8e-06;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|    :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 13
US-09-037-190-46
; Sequence 46, Application US/09037190
; Patent No. 6096515
;  GENERAL INFORMATION:
;    APPLICANT:  Crabtree, Gerald R.
;    APPLICANT:  No. 6096515throp, Jeffrey P.
;    APPLICANT:  Ho, Steffan M.
;    TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,190
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/260,174
;      FILING DATE:  13-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/124,981
;      FILING DATE:  20-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  APV-332.03
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  716 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-037-190-46

  Query Match             24.8%;  Score 114;  DB 3;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 2.8e-06;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|    :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQQPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 14
US-09-037-192-38
; Sequence 38, Application US/09037192
; Patent No. 6096860
;  GENERAL INFORMATION:
;    APPLICANT:  Crabtree, Gerald R.
;    APPLICANT:  No. 6096860throp, Jeffrey P.
;    APPLICANT:  Ho, Steffan M.
;    TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,192
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/260,174
;      FILING DATE:  13-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/124,981
;      FILING DATE:  20-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  APV-332.04
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  716 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  peptide
US-09-037-192-38

  Query Match             24.8%;  Score 114;  DB 3;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 2.8e-06;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 15
US-09-037-192-46
; Sequence 46, Application US/09037192
; Patent No. 6096860
;  GENERAL INFORMATION:
;    APPLICANT:  Crabtree, Gerald R.
;    APPLICANT:  No. 6096860throp, Jeffrey P.
;    APPLICANT:  Ho, Steffan M.
;    TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,192
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/260,174
;      FILING DATE:  13-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/124,981
;      FILING DATE:  20-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  APV-332.04
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  716 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-037-192-46

  Query Match             24.8%;  Score 114;  DB 3;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 2.8e-06;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQQPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401

Search completed: April 11, 2003, 14:24:02
Job time : 16 secs
```

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:22:26 ; Search time 14 Seconds
(without alignments)
375.550 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	114	24.8	716	9	US-10-040-430-38	Sequence 38, Appl
2	114	24.8	716	9	US-10-040-430-46	Sequence 46, Appl
3	114	24.8	716	9	US-09-349-058-38	Sequence 38, Appl
4	114	24.8	716	9	US-09-349-058-46	Sequence 46, Appl
5	65.5	14.3	557	9	US-09-925-299-940	Sequence 940, App
6	65.5	14.3	557	10	US-09-925-299-940	Sequence 940, App
7	64.5	14.1	694	9	US-09-405-920-2	Sequence 2, Appli
8	63	13.7	404	10	US-09-764-864-1244	Sequence 1244, Ap
9	62	13.5	522	10	US-09-765-272-120	Sequence 120, App
10	62	13.5	538	10	US-09-758-664-2	Sequence 2, Appli
11	62	13.5	1040	10	US-09-765-272-118	Sequence 118, App
12	61.5	13.4	427	9	US-09-965-529-20	Sequence 20, Appl
13	61.5	13.4	1368	9	US-10-159-151-2	Sequence 2, Appli
14	61	13.3	112	10	US-09-216-393-83	Sequence 83, Appl
15	61	13.3	1203	12	US-10-027-923-5	Sequence 5, Appli
16	60.5	13.2	174	9	US-09-738-626-6602	Sequence 6602, Ap
17	60	13.1	491	9	US-10-227-884-34	Sequence 34, Appl
18	60	13.1	491	9	US-10-230-163-34	Sequence 34, Appl
19	60	13.1	491	9	US-10-218-631-34	Sequence 34, Appl
20	60	13.1	491	9	US-10-230-338-34	Sequence 34, Appl
21	60	13.1	491	9	US-10-230-414-34	Sequence 34, Appl
22	60	13.1	577	10	US-09-810-264-24	Sequence 24, Appl
23	60	13.1	1213	10	US-09-969-515-8	Sequence 8, Appli
24	60	13.1	1224	10	US-09-969-515-2	Sequence 2, Appli
25	59.5	13.0	467	10	US-09-912-020-381	Sequence 381, App
26	59	12.9	218	10	US-09-825-561A-6	Sequence 6, Appli
27	59	12.9	289	9	US-09-941-973-2	Sequence 2, Appli
28	59	12.9	349	9	US-09-809-391-371	Sequence 371, App
29	59	12.9	538	9	US-09-972-218A-2	Sequence 2, Appli
30	59	12.9	538	10	US-09-965-313-2	Sequence 2, Appli
31	59	12.9	538	10	US-09-923-246-115	Sequence 115, App
32	59	12.9	538	10	US-09-825-561A-2	Sequence 2, Appli
33	59	12.9	567	10	US-09-825-561A-16	Sequence 16, Appl
34	59	12.9	606	10	US-09-923-246-97	Sequence 97, Appl
35	59	12.9	606	10	US-09-825-561A-30	Sequence 30, Appl
36	59	12.9	911	9	US-09-291-417-92	Sequence 92, Appl
37	59	12.9	3353	10	US-09-888-615-64	Sequence 64, Appl
38	58.5	12.7	592	9	US-10-227-884-62	Sequence 62, Appl
39	58.5	12.7	592	9	US-10-230-163-62	Sequence 62, Appl
40	58.5	12.7	592	9	US-10-218-631-62	Sequence 62, Appl
41	58.5	12.7	592	9	US-10-230-338-62	Sequence 62, Appl
42	58.5	12.7	592	9	US-10-230-414-62	Sequence 62, Appl
43	58.5	12.7	720	10	US-09-801-368-176	Sequence 176, App
44	58	12.6	129	10	US-09-864-761-41836	Sequence 41836, A
45	58	12.6	181	9	US-09-984-271-114	Sequence 114, App

ALIGNMENTS

RESULT 1
US-10-040-430-38
; Sequence 38, Application US/10040430
; Publication No. US20030049641A1
; GENERAL INFORMATION:
; APPLICANT: Crabtree, Gerald R.
; APPLICANT: No. US20030049641Althrop, Jeffrey P.
; APPLICANT: Ho, Steffan M.
; APPLICANT: Flanagan, William M.
; TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES AND SCREENING
; TITLE OF INVENTION: METHODS FOR IMMUNOSUPPRESSIVE AGENTS
; FILE REFERENCE: APV-008.04
; CURRENT APPLICATION NUMBER: US/10/040,430
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US/09/232,346
; PRIOR FILING DATE: 1999-01-15
; PRIOR APPLICATION NUMBER: 08/507,032
; PRIOR FILING DATE: 1995-07-31
; PRIOR APPLICATION NUMBER: 08/228,944
; PRIOR FILING DATE: 1994-04-18
; PRIOR APPLICATION NUMBER: 07/749,385
; PRIOR FILING DATE: 1991-08-22
; PRIOR APPLICATION NUMBER: 08/260,174
; PRIOR FILING DATE: 1994-06-13
; PRIOR APPLICATION NUMBER: 08/124,981
; PRIOR FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 716
; TYPE: PRT
; ORGANISM: human
US-10-040-430-38

Query Match 24.8%; Score 114; DB 9; Length 716;
Best Local Similarity 34.7%; Pred. No. 0.00011;
Matches 25; Conservative 19; Mismatches 26; Indels 2; Gaps 2;

```
Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389
```

```
Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 2
US-10-040-430-46
; Sequence 46, Application US/10040430
; Publication No. US20030049641A1
; GENERAL INFORMATION:
;  APPLICANT: Crabtree, Gerald R.
;  APPLICANT:  No. US20030049641A1throp, Jeffrey P.
;  APPLICANT:  Ho, Steffan M.
;  APPLICANT:  Flanagan, William M.
;  TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES AND SCREENING
;  TITLE OF INVENTION:  METHODS FOR IMMUNOSUPPRESSIVE AGENTS
;  FILE REFERENCE: APV-008.04
;  CURRENT APPLICATION NUMBER: US/10/040,430
;  CURRENT FILING DATE:  2002-01-07
;  PRIOR APPLICATION NUMBER: US/09/232,346
;  PRIOR FILING DATE: 1999-01-15
;  PRIOR APPLICATION NUMBER: 08/507,032
;  PRIOR FILING DATE: 1995-07-31
;  PRIOR APPLICATION NUMBER: 08/228,944
;  PRIOR FILING DATE: 1994-04-18
;  PRIOR APPLICATION NUMBER: 07/749,385
;  PRIOR FILING DATE: 1991-08-22
;  PRIOR APPLICATION NUMBER: 08/260,174
;  PRIOR FILING DATE: 1994-06-13
;  PRIOR APPLICATION NUMBER: 08/124,981
;  PRIOR FILING DATE: 1993-09-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: human
US-10-040-430-46

  Query Match             24.8%;  Score 114;  DB 9;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 0.00011;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQQPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 3
US-09-349-058-38
; Sequence 38, Application US/09349058
; Publication No. US20030056229A1
; GENERAL INFORMATION:
;  APPLICANT: Crabtree, Gerald R.
;  APPLICANT: No. US20030056229A1throp, Jeffrey P.
;  APPLICANT: Ho, Steffan N.
;  TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: APV-332.12
;  CURRENT APPLICATION NUMBER: US/09/349,058
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: 08/260,174
;  EARLIER FILING DATE: 1994-06-13
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: human
US-09-349-058-38

  Query Match             24.8%;  Score 114;  DB 9;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 0.00011;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 4
US-09-349-058-46
; Sequence 46, Application US/09349058
; Publication No. US20030056229A1
; GENERAL INFORMATION:
;  APPLICANT: Crabtree, Gerald R.
;  APPLICANT: No. US20030056229A1throp, Jeffrey P.
;  APPLICANT: Ho, Steffan N.
;  TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: APV-332.12
;  CURRENT APPLICATION NUMBER: US/09/349,058
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: 08/260,174
;  EARLIER FILING DATE: 1994-06-13
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: human
US-09-349-058-46

  Query Match             24.8%;  Score 114;  DB 9;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 0.00011;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQQPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 5
US-09-925-299-940
; Sequence 940, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 940
;   LENGTH: 557
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
```

```
;   NAME/KEY: SITE
;   LOCATION: (19)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (25)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (53)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (248)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (273)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (323)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-940

  Query Match             14.3%;  Score 65.5;  DB 9;  Length 557;
  Best Local Similarity   28.8%;  Pred. No. 23;
  Matches   23;  Conservative    7;  Mismatches   33;  Indels   17;  Gaps    3;

Qy       9 QYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ-------- 60
           ||  |  :     | |   |   |||    |  |    ||:  | :: | :
Db     446 QYVKEAKM---EEKLSNPPATCTPGKPS--SKSQNKCKPSQGLSTEENLSASITKQPIHQ 500

Qy      61 ----YPLKKDSSGDQFLSVP 76
           ;        ||   |: ||||: |
Db     501 KENIIPLLVTSNSDQFLTTP 520


RESULT 6
US-09-925-299-940
; Sequence 940, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 940
;   LENGTH: 557
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (19)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (25)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (53)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (248)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (273)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (323)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-940

  Query Match             14.3%;  Score 65.5;  DB 10;  Length 557;
  Best Local Similarity   28.8%;  Pred. No. 23;
  Matches   23;  Conservative    7;  Mismatches   33;  Indels   17;  Gaps    3;

Qy       9 QYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ-------- 60
           ||  |  :     | |   |   |||    |  |    ||:  | :: | :
Db     446 QYVKEAKM---EEKLSNPPATCTPGKPS--SKSQNKCKPSQGLSTEENLSASITKQPIHQ 500

Qy      61 ----YPLKKDSSGDQFLSVP 76
                ||   |: ||||: |
Db     501 KENIIPLLVTSNSDQFLTTP 520


RESULT 7
US-09-405-920-2
; Sequence 2, Application US/09405920
; Publication No. US20030060435A1
; GENERAL INFORMATION:
;  APPLICANT: CARILLO, Serge
;  APPLICANT: BLANCHARD, Jean-Marie
;  APPLICANT: PIECHACZYK, Marc
;  TITLE OF INVENTION: METHOD OF CANCER TREATMENT BY P53 PROTEIN CONTROL
;  FILE REFERENCE: ST94037A-US
;  CURRENT APPLICATION NUMBER: US/09/405,920
;  CURRENT FILING DATE: 1999-09-24
;  EARLIER APPLICATION NUMBER: 08/737,953
;  EARLIER FILING DATE: 1996-11-27
;  EARLIER APPLICATION NUMBER: FR94/06583
;  EARLIER FILING DATE: 1994-05-31
;  EARLIER APPLICATION NUMBER: WO PCT/FR95/00670
;  EARLIER FILING DATE: 1995-05-22
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 694
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-405-920-2

  Query Match             14.1%;  Score 64.5;  DB 9;  Length 694;
  Best Local Similarity   30.3%;  Pred. No. 39;
  Matches   23;  Conservative    9;  Mismatches   25;  Indels   19;  Gaps    4;

Qy      13 ETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLK--KDSSGD 70
           | :: |:: |   |:|     ||| | :|           |: : |:| ||   |  |
Db     290 EPELDLRSIK-EVDEAKAKEEKLEKCGED-----------DETIPSEYRLKPATDKDGK 336

Qy      71 QFLSVPSPFTWSKPKP 86
             |  |      ||||
Db     337 PLLPEPE----EKPKP 348


RESULT 8
US-09-764-864-1244
; Sequence 1244, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1244
;   LENGTH: 404
;   TYPE: PRT
;   ORGANISM: Homo sapiens
```

```
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (16)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1244

  Query Match             13.7%;  Score 63;  DB 10;  Length 404;
  Best Local Similarity   34.3%;  Pred. No. 29;
  Matches   24;  Conservative    8;  Mismatches   26;  Indels   12;  Gaps    3;

Qy      13 ETDIPLKTRKTSEDQAAILPGKLEICS---DDQGNLSPSRETSVDDGLGSQYPLKKDSSG 69
           ||  | | |:| ||         : ||      | : ||:| |     ||  ||: :|:
Db      21 ETQPPEKRRRTIEDFN-------KFCSFVLAYAGYIPPSKEESDWPASGSSSPLRGESAA 73

Qy      70 DQ--FLSVPS 77
           |   : | ||
Db      74 DSDGWDSAPS 83


RESULT 9
US-09-765-272-120
; Sequence 120, Application US/09765272
; Patent No. US20020061545A1
;   GENERAL INFORMATION:
;        APPLICANT: Choi et. al.
;        TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;        NUMBER OF SEQUENCES: 452
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/765,272
;             FILING DATE: 22-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/961,083
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Brookes, A. Anders
;             REGISTRATION NUMBER: 36,373
;             REFERENCE/DOCKET NUMBER: PB340P2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (301) 309-8504
;             TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 120:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 522 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 120:
US-09-765-272-120

  Query Match             13.5%;  Score 62;  DB 10;  Length 522;
  Best Local Similarity   32.4%;  Pred. No. 53;
  Matches   22;  Conservative    9;  Mismatches   31;  Indels    6;  Gaps    3;

Qy       9 QYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKD-- 66
           :|  : |    : |||:    :    ||  :  || :|   | | ||||  :  :  |
Db     104 KYSAQRDSQPNSTKTSD---VVHSADLE-WNQGQGKVSLQGEASGDDGLSEKSSIAADNL 159

Qy      67 SSGDQFLS 74
           || | | |
Db     160 SSNDSFAS 167


RESULT 10
US-09-758-664-2
; Sequence 2, Application US/09758664
; Patent No. US20010025022A1
; GENERAL INFORMATION:
;  APPLICANT:  KIKLY, KRISTINE KAY
;  APPLICANT:  MICHALOVICH, DAVID
;  TITLE OF INVENTION:  HNOVILR
;  FILE REFERENCE:  GP-70576-C1
;  CURRENT APPLICATION NUMBER: US/09/758,664
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER:  UK 97309517.7
;  PRIOR FILING DATE:  1997-11-26
;  PRIOR APPLICATION NUMBER:  09/187,711
;  PRIOR FILING DATE:  1998-11-06
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-758-664-2

  Query Match             13.5%;  Score 62;  DB 10;  Length 538;
  Best Local Similarity   35.0%;  Pred. No. 55;
  Matches   35;  Conservative   14;  Mismatches   33;  Indels   18;  Gaps    9;

Qy       1 LSPAPFPFQYCV--ETDIPLKTRKTSEDQA-AILPGKL--EICSDDQGN---LSPSRE-T 51
           : ||| ||    |       :   |    || |  :| |||   |:   ::|:   :|| |:
Db     118 IKPAP-PFNVTVTFSGQYNISWRSDYEDPAFYMLKGKLQYELQYRNRGDPWAVSPRRKLI 176

Qy      52 SVDDGLGSQYPL--KKDSSGD-QFLSVPSP-----FTWSK 83
           |||     | :||  :|||| : |  : | |       |||:
Db     177 SVDSRSVSLHPLEFRKDSSYELQVRAGPMPGSSYQGTWSE 216


RESULT 11
US-09-765-272-118
; Sequence 118, Application US/09765272
; Patent No. US20020061545A1
;   GENERAL INFORMATION:
;        APPLICANT: Choi et. al.
;        TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;        NUMBER OF SEQUENCES: 452
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/765,272
;             FILING DATE: 22-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/961,083
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Brookes, A. Anders
;             REGISTRATION NUMBER: 36,373
;             REFERENCE/DOCKET NUMBER: PB340P2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (301) 309-8504
;             TELEFAX: (301) 309-8512
```

```
;   INFORMATION FOR SEQ ID NO: 118:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1040 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 118:
US-09-765-272-118

  Query Match             13.5%;  Score 62;  DB 10;  Length 1040;
  Best Local Similarity   32.4%;  Pred. No. 1.3e+02;
  Matches   22;  Conservative    9;  Mismatches   31;  Indels    6;  Gaps    3;

Qy       9 QYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKD-- 66
           :|  : |    : |||:    :    ||  :  || :|   | | ||||  :  :  |
Db     104 KYSAQRDSQPNSTKTSD---VVHSADLE-WNQGQGKVSLQGEASGDDGLSEKSSIAADNL 159

Qy      67 SSGDQFLS 74
           || | | |
Db     160 SSNDSFAS 167


RESULT 12
US-09-965-529-20
; Sequence 20, Application US/09965529
; Publication No. US20020182671A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, Preeti
;  APPLICANT:  YUE, Henry
;  APPLICANT:  TANG, Y. Tom
;  APPLICANT:  BANDMAN, Olga
;  APPLICANT:  BURFORD, Neil
;  APPLICANT:  AZIMZAI, Yalda
;  APPLICANT:  BAUGHN, Mariah R.
;  APPLICANT:  LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra
;  TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0731 USA
;  CURRENT APPLICATION NUMBER: US/09/965,529
;  CURRENT FILING DATE:  2001-09-26
;  PRIOR APPLICATION NUMBER: 60/149,641; 60/164,203; PCT/US00/22315
;  PRIOR FILING DATE: 1999-08-17; 1999-11-09; 2000-08-14
;  NUMBER OF SEQ ID NOS: 74
;  SOFTWARE: PERL Program
; SEQ ID NO 20
;   LENGTH: 427
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020182671A1 1656935CD1
US-09-965-529-20

  Query Match             13.4%;  Score 61.5;  DB 9;  Length 427;
  Best Local Similarity   25.3%;  Pred. No. 46;
  Matches   20;  Conservative   13;  Mismatches   39;  Indels    7;  Gaps    2;

Qy      13 ETDIPLKTRKTSEDQAAILPGKLEICSDDQG------NLSPSRETSVDDGLGSQYPLKKD 66
           :: :|  |   | |        ||      |      :: |  ::|:||  | :| | ::
Db      21 QSRLPQWTHPNSMDNLPSAASPLEQNPSKHGAIPGGLSIGPPGKSSIDDSYG-RYDLIQN 79

Qy      67 SSGDQFLSVPSPFTWSKPK 85
           |       |  | :||: |
Db      80 SESPASPPVAVPHSWSRAK 98


RESULT 13
US-10-159-151-2
; Sequence 2, Application US/10159151
; Publication No. US20030036075A1
; GENERAL INFORMATION:
;  APPLICANT: Cytokinetics, Inc.
;  APPLICANT:  Beraud, Christopher
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
;  FILE REFERENCE: 020552-002400US
;  CURRENT APPLICATION NUMBER: US/10/159,151
;  CURRENT FILING DATE:  2002-05-31
;  PRIOR APPLICATION NUMBER: US/09/967,908
;  PRIOR FILING DATE: 2001-09-28
;  PRIOR APPLICATION NUMBER: US 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1368
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: KinI-3 motor
;   OTHER INFORMATION: domain fragment
US-10-159-151-2

  Query Match             13.4%;  Score 61.5;  DB 9;  Length 1368;
  Best Local Similarity   22.7%;  Pred. No. 2e+02;
  Matches   22;  Conservative   14;  Mismatches   22;  Indels   39;  Gaps    3;

Qy      12 VETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGL---------- 57
           |::  |:  |: : | :          ||  :| :  ||| |   |
Db     893 VDSRDPINHRRAALDHS---------CSPSKGPVDWSRENSTSSGPSPRDSLAEKPYCSQ 943

Qy      58 ----------GSQYPLKKDS-----SGDQFLSVPSP 78
                    || : |:||:     ||:    ::|||
Db     944 VDFIYRQERGGGSSFDLRKDASQSEVSGENEGNLPSP 980


RESULT 14
US-09-216-393-83
; Sequence 83, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
;   LENGTH: 112
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393-83

  Query Match             13.3%;  Score 61;  DB 10;  Length 112;
  Best Local Similarity   30.3%;  Pred. No. 9.6;
  Matches   27;  Conservative   11;  Mismatches   29;  Indels   22;  Gaps    5;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           :||:||     |::| |     :   |:|  |     |: | :||     ||     |
Db       5 VSPSPF-----VQSDSP----SSPFAQSASPPP-----SEYQDSLSLPLAESV-----SS 45

Qy      61 YPLKKDSSGDQFLSVPSPFTWSK---PKP 86
           || | :|      |||  |::   | |
Db      46 LPLAKQASPLHLTQHPSPLLWTQRASPSP 74


RESULT 15
US-10-027-923-5
```

```
; Sequence 5, Application US/10027923
; Patent No. US20020142330A1
; GENERAL INFORMATION:
;  APPLICANT: Briann Gaither Bates
;  APPLICANT:  Kamalaka Gulukota
;  APPLICANT:  Yuhong Xie
;  APPLICANT:  Janet Elizabeth Paulsen
;  TITLE OF INVENTION: NOVEL GLUTAMATE RECEPTOR MODULATORY PROTEINS AND
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: GNN-024
;  CURRENT APPLICATION NUMBER: US/10/027,923
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,589
;  PRIOR FILING DATE: 2000-12-22
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1203
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-10-027-923-5

  Query Match           13.3%;  Score 61;  DB 12;  Length 1203;
  Best Local Similarity  29.9%;  Pred. No. 2e+02;
  Matches   23;  Conservative   10;  Mismatches   38;  Indels    6;  Gaps    3;

Qy       5 PFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQG--NLSPSRETSVDDGLGSQYP 62
           |    | :  :| | |  |:  :   ||  :|:    ||   :||:||          |
Db    1088 PICSSYLIPKEIQLPTTMTTFAEIQPLPA-IEVTGGAQGATGVSPAQETPTG---AESAP 1143

Qy      63 LKKDSSGDQFLSVPSPF 79
            | |      |: ||||
Db    1144 GKPDLEELVALTPPSPF 1160


Search completed: April 11, 2003, 14:24:23
Job time : 17 secs
```

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 18, 2003, 20:50:51 ; Search time 2142 Seconds
(without alignments)
1168.460 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09550115/runat_11042003_141902_18148/app_query.fasta_1.263
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09550115_@CGN_1_1_2496_@runat_11042003_141902_18148 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	459	100.0	3602	10	MMU28807	U28807 Mus musculu
2	459	100.0	3638	10	D85612	D85612 Mouse mRNA
c 3	450	98.0	80112	2	AC128800	AC128800 Rattus no
4	450	98.0	251055	2	AC094385	AC094385 Rattus no
5	430	93.7	2406	6	I38155	I38155 Sequence 7
6	430	93.7	2406	6	I80097	I80097 Sequence 7
7	430	93.7	2647	6	I38156	I38156 Sequence 9
8	430	93.7	2647	6	I80098	I80098 Sequence 9
9	430	93.7	3583	9	HSU85430	U85430 Homo sapien
10	430	93.7	3680	9	HSU14510	U14510 Human trans
11	430	93.7	3725	9	BC001050	BC001050 Homo sapi
12	430	93.7	3769	9	HSU85429	U85429 Homo sapien
13	430	93.7	3784	9	HSU85428	U85428 Homo sapien
14	430	93.7	3969	6	AX409566	AX409566 Sequence
15	430	93.7	3969	6	I38157	I38157 Sequence 11
16	430	93.7	3969	6	I80099	I80099 Sequence 11
17	430	93.7	3969	9	HUMHFAT4A	L41067 Homo sapien
c 18	430	93.7	191565	9	HUAC004531	AC004531 Homo sapi
c 19	430	93.7	201981	2	AC130462	AC130462 Homo sapi
20	194	42.3	4594	5	AB037471	AB037471 Xenopus l
21	129	28.1	3075	10	BC028928	BC028928 Mus muscu
22	129	28.1	167462	2	AC123532	AC123532 Mus muscu
23	129	28.1	240105	2	AC098877	AC098877 Mus muscu
24	126	27.5	3637	10	AF283284	AF283284 Mus muscu
25	126	27.5	4799	10	AF309388S1	AF309388 Mus muscu
c 26	123	26.8	180194	2	AC116083	AC116083 Rattus no
27	122	26.6	2880	9	HUMNFAT3A	L41066 Homo sapien
28	122	26.6	2881	6	I38154	I38154 Sequence 5
29	122	26.6	2881	6	I80096	I80096 Sequence 5
30	122	26.6	2937	9	BC008857	BC008857 Homo sapi
c 31	122	26.6	196533	9	CNS00YVI	AL096870 Human chr
32	114	24.8	2743	6	I38153	I38153 Sequence 3
33	114	24.8	2743	6	I80095	I80095 Sequence 3
34	114	24.8	2743	9	HSU08015	U08015 Human NF-AT
35	114	24.8	2749	6	AR058939	AR058939 Sequence
36	114	24.8	2749	6	AX469982	AX469982 Sequence
37	114	24.8	2751	6	AR105239	AR105239 Sequence
38	114	24.8	2751	6	AR119160	AR119160 Sequence
39	114	24.8	2751	6	AR123531	AR123531 Sequence
40	114	24.8	2751	6	AR138184	AR138184 Sequence
41	114	24.8	2751	6	AR176745	AR176745 Sequence
42	114	24.8	2935	9	HSU59736	U59736 Human trans
43	114	24.8	4617	9	HSU80918	U80918 Homo sapien
44	114	24.8	4678	9	HSU80917	U80917 Homo sapien
45	114	24.8	5502	6	AR198316	AR198316 Sequence

ALIGNMENTS

RESULT 1

```
MMU28807
LOCUS       MMU28807                3602 bp    mRNA    linear   ROD 02-APR-1996
DEFINITION  Mus musculus lymphoid-specific transcription factor NFATc3 mRNA,
            partial cds.
ACCESSION   U28807
VERSION     U28807.1  GI:1244445
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3602)
  AUTHORS   Ho,S.N., Thomas,D.J., Timmerman,L.A., Li,X., Francke,U. and
            Crabtree,G.R.
  TITLE     NFATc3, a lymphoid-specific NFATc family member that is
            calcium-regulated and exhibits distinct DNA binding specificity
  JOURNAL   J. Biol. Chem. 270 (34), 19898-19907 (1995)
  MEDLINE   95378239
   PUBMED   7650004
REFERENCE   2  (bases 1 to 3602)
  AUTHORS   Ho,S.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1995) Steffan N. Ho, Pathology, Beckman Center
            for Molecular and Genetic Medicine Stanford, Howard Hughes Medical
            Institute, Stanford University Medical Center, 300 Pasteur Drive,
            Stanford, CA 94305, USA
COMMENT     On Apr 2, 1996 this sequence version replaced gi:968966.
FEATURES             Location/Qualifiers
     source          1. .3602
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /map="band D"
                     /clone="3"
                     /tissue_type="thymus"
                     /clone_lib="DO11.10 TCR transgenic mouse apoptotic
                     thymocyte library of B.A. Osborne et al., Nature 367,
                     281-284 (1994)"
     CDS             <1. .3198
                     /note="lymphoid-specific transcription factor"
                     /codon_start=1
                     /product="NFATc3"
                     /protein_id="AAA93249.1"
                     /db_xref="GI:1244446"
                     /translation="PRRVLFSVSAQLPSRTRPGPSDLDLEPDDCASIYIFNVDPPPST
                     LNSSLGLPHHGLLQSHSSVLSPSFQLQGYKNYEGTGDISESKYSPLGGPKPFECPSIQ
                     ITSISPNCHQGTDAHEDDLHINDPEREYLERPSRDHLYLPLEPSYRESSLSPSPASSI
                     SSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGGCPGEESWHQQ
                     YGSGHSLSPRQSPCHSPRSSITDENWLSPRPASGPSSRPTSPCGKRRHSSAEVCYAGS
                     LSPHHSPVPSPGHSPRGSVTEDTWLTAPVHTGSGLSPAPFPFQYCVETDIPLKTRKTS
                     EDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSVPSPFTWS
                     KPKPGHTPIFRTSSLPPLDWPLPTHFGQCELKIEVQPKTHHRAHYETEGSRGAVKAST
                     GGHPVVKLLGYSEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVATASQEIIIAST
                     KVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRLVFRVHIPQPS
                     GKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMIVTGSNFLPESKIIFL
                     EKGQDGRPHWEAEGKIIREKCQGAHIVLEVPPYHNPAVTSAVQVHFYLCNGKRKKSQS
                     QRFTYTPVLMKQEQKEKTQICLQFHLCLCLILPSPEALLRDRHPHDRAMSAPGSLLCQ
                     VQPAYTSMVASTHLPQLQCRDEGAGKEQHIATSSVMHQPFQVTPTSPIGSSYQSIQTS
                     MYNGPTCLPVNPASSQEFDPVLFQQDAALSSLVNLGCQPLSPIPFHSSNSDATGHLLA
                     HSPHSVQTPPHLQSMGYHCSNAGQTALSSPVADQITGQPSSHLQPITYCPSHPGSATA
                     ASPAASHPLASSPISGPSSPQLQPMPYQSPSSGTASSPSPTTRMHSGQHSTQAQSTGQ
                     GGLSVPSSLVCHSLCDPASFPPGGATVSIKPEPEDQEPNFATIGLQDITLDDVNEIIG
                     RDMSQISVSQATEVMRDTPLPGPASPDLMTSHSAQ"
     polyA_site      3602
BASE COUNT      967 a    923 c    711 g   1001 t
ORIGIN

Alignment Scores:
Pred. No.:              1.41e-42        Length:         3602
Score:                  459.00          Matches:        86
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0

DB:                     10              Gaps:           0

US-09-550-115A-2 (1-86) x MMU28807 (1-3602)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     931 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 990

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     991 AGGAAGACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGATCTGTTCAGAT 1050

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1051 GATCAAGGGAACTTATCCCCTTCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 1110

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1111 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCCTTTACC 1170

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1171 TGGAGCAAACCAAAGCCT 1188

RESULT 2
D85612
LOCUS       D85612                  3638 bp    mRNA    linear   ROD 06-FEB-1999
DEFINITION  Mouse mRNA for NFATx, complete cds.
ACCESSION   D85612
VERSION     D85612.1  GI:1906311
KEYWORDS    mNFATx1; NFATx; murine nuclear factor of activated T cell.
SOURCE      Mus musculus thymic lymphoma cultured cell cell_line:EL-4 cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Liu,J., Koyano-Nakagawa,N., Amasaki,Y., Saito-Ohara,F., Ikeuchi,T.,
            Imai,S., Takano,T., Arai,N., Yokota,T. and Arai,K.
  TITLE     Calcineurin-dependent nuclear translocation of a murine
            transcription factor NFATx: molecular cloning and functional
            characterization
  JOURNAL   Mol. Biol. Cell 8 (1), 157-170 (1997)
  MEDLINE   97170074
REFERENCE   2  (bases 1 to 3638)
  AUTHORS   Liu,J., Koyano-Nakagawa,N., Amasaki,Y., Saito-Ohara,F., Ikeuchi,T.,
            Imai,S., Takano,T., Arai,N., Yokota,T. and Arai,K.
  TITLE     Calcineurin Dependent Nuclear Translocation of a Murine
            Transcription Factor NFATx: Molecular Cloning and Functional
            Characterization
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3638)
  AUTHORS   Liu,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1996) Jie Liu, Institute of Medical Science, The
            University of Tokyo, Department of Molecular and Developmental
            Biology; 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan
            (E-mail:liujie@hgc.ims.u-tokyo.ac.jp, Tel:81-3-5449-5664,
            Fax:81-3-5449-5424)
COMMENT     On Mar 25, 1997 this sequence version replaced gi:1842164.
FEATURES             Location/Qualifiers
     source          1. .3638
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="EL-4"
                     /cell_type="cultured cell"
                     /tissue_type="thymic lymphoma"
     gene            1. .3638
                     /gene="mNFATx1"
     CDS             8. .3235
                     /gene="mNFATx1"
                     /standard_name="murine nuclear factor of activated T cell"
                     /codon_start=1
```

```
                     /product="NFATx"
                     /protein_id="BAA12833.1"
                     /db_xref="GI:1842165"
                     /translation="MTTANCGAHDELDFKLVFGEDGAPAPPPPGSRPADLEPDDCASI
                     YIFNVDPPPSTLTTPLCLPHHGLPSHSSVLSPSFQLQSHKNYEGTCEIPESKYSPLGG
                     PKPFECPSIQFTSISPNCQQELDAHEDDLQINDPEREFLERPSRDHLYLPLEPSYRES
                     SLSPSPASSISSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGG
                     CPGEESWHQQYGSGHSLSPRQSPCHSPRSSITDENWLSPRPASGPSSSRPTSPCGKRRH
                     SSAEVCYAGSLSPHHSPVPSPGHSPRGSVTEDTWLTAPVHTGSGLSPAPFPFQYCVET
                     DIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQF
                     LSVPSPFTWSKPKPGHTPIFRTSSLPPLDWPLPTHFGQCELKIEVQPKTHHRAHYETE
                     GSRGAVKASTGGHPVVKLLGYSEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVAT
                     ASQEIIIASTKVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRL
                     VFRVHIPQPSGKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMIVTGSN
                     FLPESKIIFLEKGQDGGPHWEVEGKIIREKCQGAHIVLEVPPYHNPAVTSAVQVHFYL
                     CNGKRKKSQSQRFTYTPVLMKQEQREDTDLPSVPSLPVPHSAQAQRPSSETGHPHDRA
                     MSAPGGLLCQVQPAYTSMVASTHLPQLQCRDEGAGKEQHIATSSVMHQPFQVTPTSPI
                     GSSYQSIQTSMYNGPTCLPVNPASSQEFDPVLFQQDAALSSLVNLGCQPLSPIPFHSS
                     NSDATGHLLAHSPHSVQTPPHLQSMGYHCSNAGQTALSSPVADQITGQPSSHLQPITY
                     CPSHPGSATAASPAASHPLASSPISGPSSPQLQPMPYQSPSSGTASSPSPTTRMHSGQ
                     HSTQAQSTGQGGLSVPSSLVCHSLCDPASFPPGGATVSIKPEPEDQEPNFATIGLQDI
                     TLDDVNEIIGRDMSQISVSQATEVMRDTPLPGPASPDLMTSHSAH"
     polyA_site      3638
                     /gene="mNFATx1"
                     /note="18 A nucleotides"
BASE COUNT      960 a    944 c    737 g    997 t
ORIGIN

Alignment Scores:
Pred. No.:              1.42e-42        Length:         3638
Score:                  459.00          Matches:        86
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-550-115A-2 (1-86) x D85612 (1-3638)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     968 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 1027

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1028 AGGAAGACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGATCTGTTCAGAT 1087

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1088 GATCAAGGGAACTTATCCCCTTCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 1147

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1148 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCCTTTACC 1207

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1208 TGGAGCAAACCAAAGCCT 1225

RESULT 3
AC128800/c
LOCUS       AC128800               80112 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-149I13, *** SEQUENCING IN PROGRESS
            ***, 47 unordered pieces.
ACCESSION   AC128800
VERSION     AC128800.1  GI:21952659
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 80112)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 80112)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KAMD
                Center clone name: CH230-149I13
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 45897 bases at least Q40
                Consensus quality: 49666 bases at least Q30
                Consensus quality: 52006 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 47 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
*       1     1167: contig of 1167 bp in length
*    1168     1267: gap of unknown length
*    1268     2322: contig of 1055 bp in length
*    2323     2422: gap of unknown length
*    2423     3545: contig of 1123 bp in length
*    3546     3645: gap of unknown length
*    3646     4684: contig of 1039 bp in length
*    4685     4784: gap of unknown length
*    4785     5931: contig of 1147 bp in length
*    5932     6031: gap of unknown length
*    6032     7197: contig of 1166 bp in length
*    7198     7297: gap of unknown length
*    7298     9109: contig of 1812 bp in length
*    9110     9209: gap of unknown length
*    9210    10351: contig of 1142 bp in length
*   10352    10451: gap of unknown length
*   10452    11686: contig of 1235 bp in length
*   11687    11786: gap of unknown length
*   11787    13724: contig of 1938 bp in length
*   13725    13824: gap of unknown length
*   13825    15260: contig of 1436 bp in length
*   15261    15360: gap of unknown length
*   15361    16528: contig of 1168 bp in length
*   16529    16628: gap of unknown length
*   16629    17895: contig of 1267 bp in length
*   17896    17995: gap of unknown length
*   17996    19015: contig of 1020 bp in length
*   19016    19115: gap of unknown length
*   19116    20340: contig of 1225 bp in length
*   20341    20440: gap of unknown length
*   20441    22641: contig of 2201 bp in length
*   22642    22741: gap of unknown length
*   22742    24172: contig of 1431 bp in length
*   24173    24272: gap of unknown length
*   24273    25464: contig of 1192 bp in length
*   25465    25564: gap of unknown length
*   25565    26720: contig of 1156 bp in length
*   26721    26820: gap of unknown length
*   26821    27851: contig of 1031 bp in length
*   27852    27951: gap of unknown length
*   27952    29978: contig of 2027 bp in length
*   29979    30078: gap of unknown length
*   30079    31265: contig of 1187 bp in length
*   31266    31365: gap of unknown length
*   31366    32474: contig of 1109 bp in length
*   32475    32574: gap of unknown length
*   32575    34748: contig of 2174 bp in length
*   34749    34848: gap of unknown length
*   34849    35869: contig of 1021 bp in length
*   35870    35969: gap of unknown length
*   35970    36971: contig of 1002 bp in length
*   36972    37071: gap of unknown length
*   37072    39138: contig of 2067 bp in length
*   39139    39238: gap of unknown length
*   39239    40675: contig of 1437 bp in length
*   40676    40775: gap of unknown length
*   40776    41778: contig of 1003 bp in length
*   41779    41878: gap of unknown length
*   41879    43233: contig of 1355 bp in length
*   43234    43333: gap of unknown length
*   43334    44871: contig of 1538 bp in length
*   44872    44971: gap of unknown length
*   44972    46882: contig of 1911 bp in length
*   46883    46982: gap of unknown length
*   46983    48058: contig of 1076 bp in length
*   48059    48158: gap of unknown length
*   48159    49724: contig of 1566 bp in length
*   49725    49824: gap of unknown length
*   49825    52145: contig of 2321 bp in length
*   52146    52245: gap of unknown length
*   52246    54449: contig of 2204 bp in length
*   54450    54549: gap of unknown length
*   54550    55854: contig of 1305 bp in length
*   55855    55954: gap of unknown length
*   55955    57587: contig of 1633 bp in length
*   57588    57687: gap of unknown length
*   57688    60364: contig of 2677 bp in length
*   60365    60464: gap of unknown length
*   60465    62367: contig of 1903 bp in length
*   62368    62467: gap of unknown length
*   62468    64650: contig of 2183 bp in length
*   64651    64750: gap of unknown length
*   64751    66564: contig of 1814 bp in length
*   66565    66664: gap of unknown length
*   66665    69184: contig of 2520 bp in length
*   69185    69284: gap of unknown length
*   69285    71426: contig of 2142 bp in length
*   71427    71526: gap of unknown length
*   71527    75215: contig of 3689 bp in length
*   75216    75315: gap of unknown length
*   75316    77221: contig of 1906 bp in length
*   77222    77321: gap of unknown length
*   77322    80112: contig of 2791 bp in length.
FEATURES             Location/Qualifiers
     source          1. .80112
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-149I13"
BASE COUNT    18233 a   18609 c   21607 g   16798 t    4865 others
ORIGIN

Alignment Scores:
Pred. No.:                3.52e-40        Length:          80112
Score:                    450.00          Matches:         83
Percent Similarity:       100.00%         Conservative:    3
Best Local Similarity:    96.51%          Mismatches:      0
Query Match:              98.04%          Indels:          0
DB:                       2               Gaps:            0

US-09-550-115A-2 (1-86) x AC128800 (1-80112)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38499 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 38440

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           |||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db   38439 AGGAAGACTTCTGACGATCAAGCTGCCATACTACCAGGAAAATTAGAGGTCTGTTCAGAT 38380

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38379 GATCAAGGGAGCTTATCCCCATCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 38320

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38319 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCTTTTACC 38260

Qy      81 TrpSerLysProLysPro 86
           ||||||||||||||||||
Db   38259 TGGAGCAAACCAAAGCCT 38242

RESULT 4
AC094385
LOCUS       AC094385              251055 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-3N11, *** SEQUENCING IN PROGRESS ***,
            52 unordered pieces.
ACCESSION   AC094385
VERSION     AC094385.4  GI:21716525
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 251055)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
```

```
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Bärbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 251055)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 251055)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 9, 2002 this sequence version replaced gi:17941108.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAMP
                Center clone name: CH230-3N11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 204355 bases at least Q40
                Consensus quality: 208911 bases at least Q30
                Consensus quality: 212278 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 52 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1108: contig of 1108 bp in length
            *    1109     1208: gap of unknown length
            *    1209     2683: contig of 1475 bp in length
            *    2684     2783: gap of unknown length
            *    2784     3883: contig of 1100 bp in length
            *    3884     3983: gap of unknown length
            *    3984     5519: contig of 1536 bp in length
            *    5520     5619: gap of unknown length
            *    5620     7019: contig of 1400 bp in length
            *    7020     7119: gap of unknown length
            *    7120     8194: contig of 1075 bp in length
            *    8195     8294: gap of unknown length
            *    8295     9646: contig of 1352 bp in length
            *    9647     9746: gap of unknown length
            *    9747    10988: contig of 1242 bp in length
            *   10989    11088: gap of unknown length
            *   11089    12378: contig of 1290 bp in length
            *   12379    12478: gap of unknown length
            *   12479    13746: contig of 1268 bp in length
            *   13747    13846: gap of unknown length
            *   13847    15479: contig of 1633 bp in length
            *   15480    15579: gap of unknown length
            *   15580    17146: contig of 1567 bp in length
            *   17147    17246: gap of unknown length
            *   17247    18699: contig of 1453 bp in length
            *   18700    18799: gap of unknown length
            *   18800    20609: contig of 1810 bp in length
            *   20610    20709: gap of unknown length
            *   20710    22107: contig of 1398 bp in length
            *   22108    22207: gap of unknown length
            *   22208    23709: contig of 1502 bp in length
            *   23710    23809: gap of unknown length
            *   23810    24812: contig of 1003 bp in length
            *   24813    24912: gap of unknown length
            *   24913    27254: contig of 2342 bp in length
            *   27255    27354: gap of unknown length
            *   27355    29236: contig of 1882 bp in length
            *   29237    29336: gap of unknown length
            *   29337    31709: contig of 2373 bp in length
            *   31710    31809: gap of unknown length
            *   31810    33833: contig of 2024 bp in length
            *   33834    33933: gap of unknown length
            *   33934    36015: contig of 2082 bp in length
            *   36016    36115: gap of unknown length
            *   36116    38441: contig of 2326 bp in length
            *   38442    38541: gap of unknown length
            *   38542    42767: contig of 4226 bp in length
            *   42768    42867: gap of unknown length
            *   42868    44632: contig of 1765 bp in length
            *   44633    44732: gap of unknown length
            *   44733    47340: contig of 2608 bp in length
            *   47341    47440: gap of unknown length
            *   47441    49274: contig of 1834 bp in length
            *   49275    49374: gap of unknown length
            *   49375    54632: contig of 5258 bp in length
            *   54633    54732: gap of unknown length
            *   54733    58371: contig of 3639 bp in length
            *   58372    58471: gap of unknown length
            *   58472    61803: contig of 3332 bp in length
            *   61804    61903: gap of unknown length
            *   61904    66184: contig of 4281 bp in length
            *   66185    66284: gap of unknown length
            *   66285    69862: contig of 3578 bp in length
            *   69863    69962: gap of unknown length
            *   69963    75824: contig of 5862 bp in length
```

```
            *    75825    75924: gap of unknown length
            *    75925    81868: contig of 5944 bp in length
            *    81869    81968: gap of unknown length
            *    81969    87828: contig of 5860 bp in length
            *    87829    87928: gap of unknown length
            *    87929    94012: contig of 6084 bp in length
            *    94013    94112: gap of unknown length
            *    94113    99920: contig of 5808 bp in length
            *    99921   100020: gap of unknown length
            *   100021   105508: contig of 5488 bp in length
            *   105509   105608: gap of unknown length
            *   105609   111533: contig of 5925 bp in length
            *   111534   111633: gap of unknown length
            *   111634   117749: contig of 6116 bp in length
            *   117750   117849: gap of unknown length
            *   117850   125485: contig of 7636 bp in length
            *   125486   125585: gap of unknown length
            *   125586   132663: contig of 7078 bp in length
            *   132664   132763: gap of unknown length
            *   132764   139566: contig of 6803 bp in length
            *   139567   139666: gap of unknown length
            *   139667   149287: contig of 9621 bp in length
            *   149288   149387: gap of unknown length
            *   149388   159110: contig of 9723 bp in length
            *   159111   159210: gap of unknown length
            *   159211   166473: contig of 7263 bp in length
            *   166474   166573: gap of unknown length
            *   166574   177703: contig of 11130 bp in length
            *   177704   177803: gap of unknown length
            *   177804   188488: contig of 10685 bp in length
            *   188489   188588: gap of unknown length
            *   188589   199293: contig of 10705 bp in length
            *   199294   199393: gap of unknown length
            *   199394   211045: contig of 11652 bp in length
            *   211046   211145: gap of unknown length
            *   211146   230441: contig of 19296 bp in length
            *   230442   230541: gap of unknown length
            *   230542   251055: contig of 20514 bp in length.
FEATURES             Location/Qualifiers
     source          1. .251055
                     /organism="Rattus norvegicus"

Alignment Scores:
Pred. No.:              1.13e-39       Length:         251055
Score:                  450.00         Matches:        83
Percent Similarity:     100.00%        Conservative:   3
Best Local Similarity:  96.51%         Mismatches:     0
Query Match:            98.04%         Indels:         0
DB:                     2              Gaps:           0

US-09-550-115A-2 (1-86) x AC094385 (1-251055)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46000 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 46059

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||:::||||||||||||||||||||||||||||||||:::|||||||||
Db  46060 AGGAAGACTTCTGACGATCAAGCTGCCATACTACCAGGAAAATTAGAGGTCTGTTCAGAT 46119

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46120 GATCAAGGGAGCTTATCCCCATCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 46179

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  46180 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCTTTTACC 46239

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db  46240 TGGAGCAAACCAAAGCCT 46257

RESULT 5
I38155
LOCUS       I38155                 2406 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 7 from patent US 5612455.
ACCESSION   I38155
VERSION     I38155.1  GI:2086145
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2406)
  AUTHORS   Hoey,T.
  TITLE     Nuclear factors and binding assay
  JOURNAL   Patent: US 5612455-A 7 18-MAR-1997;
FEATURES             Location/Qualifiers
     source          1. .2406
                     /organism="unknown"
BASE COUNT      645 a    612 c    535 g    614 t
ORIGIN

Alignment Scores:
Pred. No.:              1.89e-39       Length:         2406
Score:                  430.00         Matches:        79
Percent Similarity:     96.51%         Conservative:   4
Best Local Similarity:  91.86%         Mismatches:     3
Query Match:            93.68%         Indels:         0
DB:                     6              Gaps:           0

US-09-550-115A-2 (1-86) x I38155 (1-2406)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::||||||||||:::||||||||||||||:::||||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428

RESULT 6
I80097
LOCUS       I80097                 2406 bp    DNA     linear   PAT 10-JUN-1998
DEFINITION  Sequence 7 from patent US 5708158.
ACCESSION   I80097
VERSION     I80097.1  GI:3208387
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2406)
  AUTHORS   Hoey,T.
  TITLE     Nuclear factors and binding assays
  JOURNAL   Patent: US 5708158-A 7 13-JAN-1998;
FEATURES             Location/Qualifiers
     source          1. .2406
                     /organism="unknown"
BASE COUNT      645 a    612 c    535 g    614 t
ORIGIN

Alignment Scores:
Pred. No.:              1.89e-39       Length:         2406
Score:                  430.00         Matches:        79
Percent Similarity:     96.51%         Conservative:   4
```

```
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-550-115A-2 (1-86) x I80097 (1-2406)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 7
I38156
LOCUS       I38156                  2647 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 9 from patent US 5612455.
ACCESSION   I38156
VERSION     I38156.1  GI:2086146
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2647)
  AUTHORS   Hoey,T.
  TITLE     Nuclear factors and binding assay
  JOURNAL   Patent: US 5612455-A 9 18-MAR-1997;
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="unknown"
BASE COUNT      656 a    711 c    584 g    696 t
ORIGIN

Alignment Scores:
Pred. No.:              2.08e-39        Length:         2647
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-550-115A-2 (1-86) x I38156 (1-2647)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428
```

```
RESULT 8
I80098
LOCUS       I80098                  2647 bp    DNA     linear   PAT 10-JUN-1998
DEFINITION  Sequence 9 from patent US 5708158.
ACCESSION   I80098
VERSION     I80098.1  GI:3208388
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2647)
  AUTHORS   Hoey,T.
  TITLE     Nuclear factors and binding assays
  JOURNAL   Patent: US 5708158-A 9 13-JAN-1998;
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="unknown"
BASE COUNT      656 a    711 c    584 g    696 t
ORIGIN

Alignment Scores:
Pred. No.:              2.08e-39        Length:         2647
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-550-115A-2 (1-86) x I80098 (1-2647)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 9
HSU85430
LOCUS       HSU85430                3583 bp    mRNA    linear   PRI 03-MAY-2001
DEFINITION  Homo sapiens transcription factor NFATx4 mRNA, complete cds.
ACCESSION   U85430
VERSION     U85430.1  GI:1835590
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3583)
  AUTHORS   Masuda,E.S., Naito,Y., Tokumitsu,H., Campbell,D., Saito,F.,
            Hannum,C., Arai,K. and Arai,N.
  TITLE     NFATx, a novel member of the nuclear factor of activated T cells
            family that is expressed predominantly in the thymus
  JOURNAL   Mol. Cell. Biol. 15 (5), 2697-2706 (1995)
  MEDLINE   95257951
   PUBMED   7739550
REFERENCE   2  (bases 1 to 3583)
```

```
  AUTHORS   Imamura,R., Masuda,E.S., Naito,Y., Imai,S., Fujino,T., Takano,T.,
            Arai,K. and Arai,N.
  TITLE     Carboxyl-terminal 15-amino acid sequence of NFATx1 is possibly
            created by tissue-specific splicing and is essential for
            transactivation activity in T cells
  JOURNAL   J. Immunol. 161 (7), 3455-3463 (1998)
  MEDLINE   98430664
   PUBMED   9759864
REFERENCE   3  (bases 1 to 3583)
  AUTHORS   Arai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1997) Cell Signaling, DNAX Research Institute,
            901 California Ave, Palo Alto, CA 94304-1104, USA
FEATURES             Location/Qualifiers
     source          1. .3583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="T antigen-transformed diploid fibroblast"
                     /cell_type="fibroblast"
     CDS             25. .3162
                     /note="isoform of NFATx"
                     /codon_start=1
                     /product="transcription factor NFATx4"
                     /protein_id="AAB46597.1"
                     /db_xref="GI:1835591"
                     /translation="MTTANCGAHDELDFKLVFGEDGAPAPPPPGSRPADLEPDDCASI
                     YIFNVDPPPSTLTTPLCLPHHGLPSHSSVLSPSFQLQSHKNYEGTCEIPESKYSPLGG
                     PKPFECPSIQITSISPNCHQELDAHEDDLQINDPEREFLERPSRDHLYLPLEPSYRES
                     SLSPSPASSISSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGG
                     CPGEETWHQQYGLGHSLSPRQSPCHSPRSSVTDENWLSPRPASGPSSRPTSPCGKRRH
                     SSAEVCYAGSLSPHHSPVPSPGHSPRGSVTEDTWLNASVHGGSGLGPAVFPFQYCVET
                     DIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQYPLKKDSCGDQF
                     LSVPSPFTWSKPKPGHTPIFRTSSLPPLDWPLPAHFGQCELKIEVQPKTHHRAHYETE
                     GSRGAVKASTGGHPVVKLLGYNEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVAT
                     ASQEIIIASTKVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRL
                     VFRVHIPQPSGKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMVVTGSN
                     FLPESKIIFLEKGQDGRPQWEVEGKIIREKCQGAHIVLEVPPYHNPAVTAAVQVHFYL
                     CNGKRKKSQSQRFTYTPVLLKQEHREEIDLSSVPSLPVPHPAQTQRPSSDSGCSHDSV
                     LSGQRSLICSIPQTYASMVTSSHLPQLQCRDESVSKEQHMIPSPIVHQPFQVTPTPPV
                     GSSYQPMQTNVVYNGPTCLPINAASSQEFDSVWFQQDATLSGLVNLGCQPLSSIPFHS
                     SNSGSTGHLLAHTPHSVHTLPHLQSMGYHCSNTGQRSLSSPVGDQITGQPSSQLQPIT
                     YGPSHSGSVTTASPAASHPLGSSPLSGPPSPQFQPMPYQSPSSGTGSSPSPATRMHSG
                     QHSTQAQSTGQGGLSAPSSLICHSLCDPASFPPDGATVSIKPEPEDREPNFATIGLQD
                     ITLDDGKFISDMFLK"
BASE COUNT      959 a    898 c    757 g    969 t
ORIGIN

Alignment Scores:
Pred. No.:              2.83e-39        Length:         3583
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     9               Gaps:           0

US-09-550-115A-2 (1-86) x HSU85430 (1-3583)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::||||||||:::||||||||||||:::||||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242

RESULT 10
HSU14510
LOCUS       HSU14510               3680 bp    mRNA    linear   PRI 31-JAN-1996
DEFINITION  Human transcription factor NFATx mRNA, complete cds.
ACCESSION   U14510
VERSION     U14510.1  GI:780373
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3680)
  AUTHORS   Masuda,E.S., Naito,Y., Tokumitsu,H., Campbell,D., Saito,F.,
            Hannum,C., Arai,K. and Arai,N.
  TITLE     NFATx, a novel member of the nuclear factor of activated T cells
            family that is expressed predominantly in the thymus
  JOURNAL   Mol. Cell. Biol. 15 (5), 2697-2706 (1995)
  MEDLINE   95257951
   PUBMED   7739550
REFERENCE   2  (bases 1 to 3680)
  AUTHORS   Arai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1994) Naoko Arai, Molecular Biology, DNAX
            Research Institute of Molecular and Cellular Biology, 901
            California Avenue, Palo Alto, CA 94304-1104, USA
FEATURES             Location/Qualifiers
     source          1. .3680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q21-q22"
                     /cell_line="Jurkat"
                     /cell_type="T-cell"
                     /note="in LambdaZipox vector"
     CDS             25. .3252
                     /function="transcription factor"
                     /codon_start=1
                     /product="NFATx"
                     /protein_id="AAA86308.1"
                     /db_xref="GI:780374"
                     /translation="MTTANCGAHDELDFKLVFGEDGAPAPPPPGSRPADLEPDDCASI
                     YIFNVDPPPSTLTTPLCLPHHGLPSHSSVLSPSFQLQSHKNYEGTCEIPESKYSPLGG
                     PKPFECPSIQITSISPNCHQELDAHEDDLQINDPEREFLERPSRDHLYLPLEPSYRES
                     SLSPSPASSISSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGG
                     CPGEETWHQQYGLGHSLSPRQSPCHSPRSSVTDENWLSPRPASGPSSRPTSPCGKRRH
                     SSAEVCYAGSLSPHHSPVPSPGHSPRGSVTEDTWLNASVHGGSGLGPAVFPFQYCVET
                     DIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQYPLKKDSCGDQF
                     LSVPSPFTWSKPKPGHTPIFRTSSLPPLDWPLPAHFGQCELKIEVQPKTHHRAHYETE
                     GSRGAVKASTGGHPVVKLLGYNEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVAT
                     ASQEIIIASTKVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRL
                     VFRVHIPQPSGKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMVVTGSN
                     FLPESKIIFLEKGQDGRPQWEVEGKIIREKCQGAHIVLEVPPYHNPAVTAAVQVHFYL
                     CNGKRKKSQSQRFTYTPVLMKQEHREEIDLSSVPSLPVPHPAQTQRPSSDSGCSHDSV
                     LSGQRSLICSIPQTYASMVTSSHLPQLQCRDESVSKEQHMIPSPIVHQPFQVTPTPPV
                     GSSYQPMQTNVVYNGPTCLPINAASSQEFDSVLFQQDATLSGLVNLGCQPLSSIPFHS
                     SNSGSTGHLLAHTPHSVHTLPHLQSMGYHCSNTGQRSLSSPVADQITGQPSSQLQPIT
                     YGPSHSGSATTASPAASHPLASSPLSGPPSPQLQPMPYQSPSSGTASSPSPATRMHSG
                     QHSTQAQSTGQGGLSAPSSLICHSLCDPASFPPDGATVSIKPEPEDREPNFATIGLQD
                     ITLDDVNEIIGRDMSQISVSQGAGVSRQAPLPSPESLDLGRSDGL"
BASE COUNT      948 a    926 c    781 g   1025 t
ORIGIN

Alignment Scores:
Pred. No.:              2.91e-39        Length:         3680
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     9               Gaps:           0
```

```
US-09-550-115A-2 (1-86) x HSU14510 (1-3680)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242

RESULT 11
BC001050
LOCUS       BC001050               3725 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, nuclear factor of activated T-cells, cytoplasmic,
            calcineurin-dependent 3, clone MGC:1495 IMAGE:3505967, mRNA,
            complete cds.
ACCESSION   BC001050
VERSION     BC001050.1  GI:12654448
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3725)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: d Column: 19
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4758803.
FEATURES             Location/Qualifiers
     source          1. .3725
                     /organism="Homo sapiens"
                     /db_xref="LocusID:4775"
                     /db_xref="taxon:9606"
                     /clone="MGC:1495 IMAGE:3505967"
                     /tissue_type="Kidney, renal cell adenocarcinoma"
                     /clone_lib="NIH_MGC_14"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             44. .3271
                     /codon_start=1
                     /product="nuclear factor of activated T-cells,
                     cytoplasmic, calcineurin-dependent 3"
                     /protein_id="AAH01050.1"
                     /db_xref="GI:12654449"
                     /translation="MTTANCGAHDELDFKLVFGEDGAPAPPPPGSRPADLEPDDCASI
                     YIFNVDPPPSTLTTPLCLPHHGLPSHSSVLSPSFQLQSHKNYEGTCEIPESKYSPLGG
                     PKPFECPSIQITSISPNCHQELDAHEDDLQINDPEREFLERPSRDHLYLPLEPSYRES
                     SLSPSPASSISSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGG
                     CPGEETWHQQYGLGHSLSPRQSPCHSPRSSVTDENWLSPRPASGPSSSRPTSPCGKRRH
                     SSAEVCYAGSLSPHHSPVPSPGHSPRGSVTEDTWLNASVHGGSGLGPAVFPFQYCVET
                     DIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQYPLKKDSCGDQF
                     LSVPSPFTWSKPKPGHTPIFRTSSLPPLDWPLPAHFGQCELKIEVQPKTHHRAHYETE
                     GSRGAVKASTGGHPVVKLLGYNEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVAT
                     ASQEIIIASTKVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRL
                     VFRVHIPQPSGKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMVVTGSN
                     FLPESKIIFLEKGQDGRPQWEVEGKIIREKCQGAHIVLEVPPYHNPAVTAAVQVHFYL
                     CNGKRKKSQSQRFTYTPVLMKQEHREEIDLSSVPSLPVPHPAQTQRPSSDSGCSHDSV
                     LSGQRSLICSIPQTYASMVTSSHLPQLQCRDESVSKEQHMIPSPIVHQPFQVTPTPPV
                     GSSYQPMQTNVVYNGPTCLPINAASSQEFDSVLFQQDATLSGLVNLGCQPLSSIPFHS
                     SNSGSTGHLLAHTPHSVHTLPHLQSMGYHCSNTGQRSLSSPVADQITGQPSSQLQPIT
                     YGPSHSGSATTASPAASHPLASSPLSGPPSPQLQPMPYQSPSSGTASSPSPATRMHSG
                     QHSTQAQSTGQGGLSAPSSLICHSLCDPASFPPDGATVSIKPEPEDREPNFATIGLQD
                     ITLDDVNEIIGRDMSQISVSQGAGVSRQAPLPSPESLDLGRSDGL"
BASE COUNT      979 a    929 c    791 g   1026 t
ORIGIN

Alignment Scores:
Pred. No.:              2.95e-39        Length:         3725
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     9               Gaps:           0

US-09-550-115A-2 (1-86) x BC001050 (1-3725)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1004 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1063

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1064 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1123

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1124 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1183

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1184 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1243

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1244 TGGAGCAAACCAAAGCCT 1261

RESULT 12
HSU85429
LOCUS       HSU85429               3769 bp    mRNA    linear   PRI 03-MAY-2001
DEFINITION  Homo sapiens transcription factor NFATx3 mRNA, complete cds.
ACCESSION   U85429
VERSION     U85429.1  GI:1835588
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3769)
  AUTHORS   Masuda,E.S., Naito,Y., Tokumitsu,H., Campbell,D., Saito,F.,
            Hannum,C., Arai,K. and Arai,N.
  TITLE     NFATx, a novel member of the nuclear factor of activated T cells
            family that is expressed predominantly in the thymus
  JOURNAL   Mol. Cell. Biol. 15 (5), 2697-2706 (1995)
  MEDLINE   95257951
```

```
   PUBMED   7739550
REFERENCE   2  (bases 1 to 3769)
  AUTHORS   Imamura,R., Masuda,E.S., Naito,Y., Imai,S., Fujino,T., Takano,T.,
            Arai,K. and Arai,N.
  TITLE     Carboxyl-terminal 15-amino acid sequence of NFATx1 is possibly
            created by tissue-specific splicing and is essential for
            transactivation activity in T cells
  JOURNAL   J. Immunol. 161 (7), 3455-3463 (1998)
  MEDLINE   98430664
   PUBMED   9759864
REFERENCE   3  (bases 1 to 3769)
  AUTHORS   Arai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1997) Cell Signaling, DNAX Research Institute,
            901 California Ave, Palo Alto, CA 94304-1104, USA
FEATURES             Location/Qualifiers
     source          1. .3769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Jurkat"
                     /cell_type="T-cell"
     CDS             25. .3222
                     /note="isoform of NFATx"
                     /codon_start=1
                     /product="transcription factor NFATx3"
                     /protein_id="AAB46596.1"
                     /db_xref="GI:1835589"
                     /translation="MTTANCGAHDELDFKLVFGEDGAPAPPPPGSRPADLEPDDCASI
                     YIFNVDPPPSTLTTPLCLPHHGLPSHSSVLSPSFQLQSHKNYEGTCEIPESKYSPLGG
                     PKPFECPSIQITSISPNCHQELDAHEDDLQINDPEREFLERPSRDHLYLPLEPSYRES
                     SLSPSPASSISSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGG
                     CPGEETWHQQYGLGHSLSPRQSPCHSPRSSVTDENWLSPRPASGPSSRPTSPCGKRRH
                     SSAEVCYAGSLSPHHSPVPSPGHSPRGSVTEDTWLNASVHGGSGLGPAVFPFQYCVET
                     DIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQYPLKKDSCGDQF
                     LSVPSPFTWSKPKPGHTPIFRTSSLPPLDWPLPAHFGQCELKIEVQPKTHHRAHYETE
                     GSRGAVKASTGGHPVVKLLGYNEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVAT
                     ASQEIIIASTKVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRL
                     VFRVHIPQPSGKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMVVTGSN
                     FLPESKIIFLEKGQDGRPQWEVEGKIIREKCQGAHIVLEVPPYHNPAVTAAVQVHFYL
                     CNGKRKKSQSQRFTYTPVLLKQEHREEIDLSSVPSLPVPHPAQTQRPSSDSGCSHDSV
                     LSGQRSLICSIPQTYASMVTSSHLPQLQCRDESVSKEQHMIPSPIVHQPFQVTPTPPV
                     GSSYQPMQTNVVYNGPTCLPINAASSQEFDSVWFQQDATLSGLVNLGCQPLSSIPFHS
                     SNSGSTGHLLAHTPHSVHTLPHLQSMGYHCSNTGQRSLSSPVGDQITGQPSSQLQPIT
                     YGPSHSGSVTTASPAASHPLGSSPLSGPPSPQFQPMPYQSPSSGTGSSPSPATRMHSG
                     QHSTQAQSTGQGGLSAPSSLICHSLCDPASFPPDGATVSIKPEPEDREPNFATIGLQD
                     ITLDDDLFTSNNFDLLQLRPTFWPVPAGRYLRNLE"
BASE COUNT      974 a    939 c    802 g   1054 t
ORIGIN

Alignment Scores:
Pred. No.:                2.98e-39       Length:         3769
Score:                    430.00         Matches:        79
Percent Similarity:       96.51%         Conservative:   4
Best Local Similarity:    91.86%         Mismatches:     3
Query Match:              93.68%         Indels:         0
DB:                       9              Gaps:           0

US-09-550-115A-2 (1-86) x HSU85429 (1-3769)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||||:::|||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||

Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242

RESULT 13
HSU85428
LOCUS       HSU85428                3784 bp    mRNA    linear   PRI 03-MAY-2001
DEFINITION  Homo sapiens transcription factor NFATx2 mRNA, complete cds.
ACCESSION   U85428
VERSION     U85428.1  GI:1835586
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3784)
  AUTHORS   Masuda,E.S., Naito,Y., Tokumitsu,H., Campbell,D., Saito,F.,
            Hannum,C., Arai,K. and Arai,N.
  TITLE     NFATx, a novel member of the nuclear factor of activated T cells
            family that is expressed predominantly in the thymus
  JOURNAL   Mol. Cell. Biol. 15 (5), 2697-2706 (1995)
  MEDLINE   95257951
   PUBMED   7739550
REFERENCE   2  (bases 1 to 3784)
  AUTHORS   Imamura,R., Masuda,E.S., Naito,Y., Imai,S., Fujino,T., Takano,T.,
            Arai,K. and Arai,N.
  TITLE     Carboxyl-terminal 15-amino acid sequence of NFATx1 is possibly
            created by tissue-specific splicing and is essential for
            transactivation activity in T cells
  JOURNAL   J. Immunol. 161 (7), 3455-3463 (1998)
  MEDLINE   98430664
   PUBMED   9759864
REFERENCE   3  (bases 1 to 3784)
  AUTHORS   Arai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-1997) Cell Signaling, DNAX Research Institute,
            901 California Ave, Palo Alto, CA 94304-1104, USA
FEATURES             Location/Qualifiers
     source          1. .3784
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Jurkat"
                     /cell_type="T-cell"
     CDS             25. .3231
                     /note="isoform of NFATx"
                     /codon_start=1
                     /product="transcription factor NFATx2"
                     /protein_id="AAB46595.1"
                     /db_xref="GI:1835587"
                     /translation="MTTANCGAHDELDFKLVFGEDGAPAPPPPGSRPADLEPDDCASI
                     YIFNVDPPPSTLTTPLCLPHHGLPSHSSVLSPSFQLQSHKNYEGTCEIPESKYSPLGG
                     PKPFECPSIQITSISPNCHQELDAHEDDLQINDPEREFLERPSRDHLYLPLEPSYRES
                     SLSPSPASSISSRSWFSDASSCESLSHIYDDVDSELNEAAARFTLGSPLTSPGGSPGG
                     CPGEETWHQQYGLGHSLSPRQSPCHSPRSSVTDENWLSPRPASGPSSRPTSPCGKRRH
                     SSAEVCYAGSLSPHHSPVPSPGHSPRGSVTEDTWLNASVHGGSGLGPAVFPFQYCVET
                     DIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQYPLKKDSCGDQF
                     LSVPSPFTWSKPKPGHTPIFRTSSLPPLDWPLPAHFGQCELKIEVQPKTHHRAHYETE
                     GSRGAVKASTGGHPVVKLLGYNEKPINLQMFIGTADDRYLRPHAFYQVHRITGKTVAT
                     ASQEIIIASTKVLEIPLLPENNMSASIDCAGILKLRNSDIELRKGETDIGRKNTRVRL
                     VFRVHIPQPSGKVLSLQIASIPVECSQRSAQELPHIEKYSINSCSVNGGHEMVVTGSN
                     FLPESKIIFLEKGQDGRPQWEVEGKIIREKCQGAHIVLEVPPYHNPAVTAAVQVHFYL
                     CNGKRKKSQSQRFTYTPVLLKQEHREEIDLSSVPSLPVPHPAQTQRPSSDSGCSHDSV
                     LSGQRSLICSIPQTYASMVTSSHLPQLQCRDESVSKEQHMIPSPIVHQPFQVTPTPPV
                     GSSYQPMQTNVVYNGPTCLPINAASSQEFDSVWFQQDATLSGLVNLGCQPLSSIPFHS
                     SNSGSTGHLLAHTPHSVHTLPHLQSMGYHCSNTGQRSLSSPVGDQITGQPSSQLQPIT
                     YGPSHSGSVTTASPAASHPLGSSPLSGPPSPQFQPMPYQSPSSGTGSSPSPATRMHSG
                     QHSTQAQSTGQGGLSAPSSLICHSLCDPASFPPDGATVSIKPEPEDREPNFATIGLQD
                     ITLDDDQFISDLEHQPSGSAEKWPNHSVLSCPAPFWRI"
BASE COUNT      979 a    950 c    804 g   1051 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.99e-39        Length:         3784
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     9               Gaps:           0

US-09-550-115A-2 (1-86) x HSU85428 (1-3784)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::|||||||||:::|||||||||:::||||||||||||||||||||||
Db   1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1225 TGGAGCAAACCAAAGCCT 1242

RESULT 14
AX409566
LOCUS       AX409566                3969 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2213 from Patent WO0229103.
ACCESSION   AX409566
VERSION     AX409566.1  GI:21442271
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2213 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .3969
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. L41067"
BASE COUNT      993 a   1021 c    875 g   1080 t
ORIGIN

Alignment Scores:
Pred. No.:              3.14e-39        Length:         3969
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-550-115A-2 (1-86) x AX409566 (1-3969)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::|||||||||:::|||||||||:::||||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428

RESULT 15
I38157
LOCUS       I38157                  3969 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 11 from patent US 5612455.
ACCESSION   I38157
VERSION     I38157.1  GI:2086147
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3969)
  AUTHORS   Hoey,T.
  TITLE     Nuclear factors and binding assay
  JOURNAL   Patent: US 5612455-A 11 18-MAR-1997;
FEATURES             Location/Qualifiers
     source          1. .3969
                     /organism="unknown"
BASE COUNT      993 a   1021 c    875 g   1080 t
ORIGIN

Alignment Scores:
Pred. No.:              3.14e-39        Length:         3969
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-550-115A-2 (1-86) x I38157 (1-3969)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::|||||||||:::|||||||||:::||||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428


Search completed: April 18, 2003, 23:01:06
Job time : 2211 secs
```

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 18, 2003, 20:52:51 ; Search time 235 Seconds
(without alignments)
824.136 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09550115/runat_11042003_141902_18138/app_query.fasta_1.263
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09550115_@CGN_1_1_263_@runat_11042003_141902_18138 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEQUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	430	93.7	2406	17	AAT33677	Human transcriptio
	2	430	93.7	3583	21	AAF21183	Human low adenosin
	3	430	93.7	3583	21	AAA35061	Human adenosine re
	4	430	93.7	3769	21	AAF21182	Human low adenosin
	5	430	93.7	3769	21	AAA35060	Human adenosine re
	6	430	93.7	3784	21	AAF21181	Human low adenosin
	7	430	93.7	3784	21	AAA35059	Human adenosine re
	8	430	93.7	3969	24	ABN95715	Gene #2213 used to
	9	430	93.7	4010	16	AAQ84303	Human NF-AT120, X
	10	430	93.7	17069	21	AAF21185	Human low adenosin
	11	430	93.7	17069	21	AAA35063	Human adenosine re
c	12	187	40.7	240	21	AAC14360	Human secreted pro
	13	122	26.6	2881	17	AAT36868	Human transcriptio
	14	122	26.6	2881	20	AAX08717	NF-AT3 encoding se
	15	114	24.8	2743	17	AAT36867	Human transcriptio
	16	114	24.8	2743	22	AAH02897	Human shear stress
	17	114	24.8	2749	16	AAQ86687	Human NF-AT transc
	18	114	24.8	2750	21	AAA29252	Human nuclear fact
	19	114	24.8	2751	22	AAF31710	Human NF-ATc cDNA.
	20	114	24.8	2751	24	AAD34048	Human cytoplasmic
	21	114	24.8	2853	16	AAQ84301	Human NF-AT120, C
	22	114	24.8	2935	21	AAF21184	Human low adenosin
	23	114	24.8	2935	21	AAA35062	Human adenosine re
	24	114	24.8	4963	23	AAS78396	DNA encoding novel
	25	114	24.8	5502	24	AAD34036	Human cytoplasmic
c	26	79.5	17.3	4370	24	ABK84613	Human cDNA differe
	27	77	16.8	4064	23	ABL20913	Drosophila melanog
	28	77	16.8	4803	23	ABL16961	Drosophila melanog
	29	77	16.8	6225	23	ABL20912	Drosophila melanog
	30	77	16.8	7027	23	ABL16960	Drosophila melanog
	31	76	16.6	32768	20	AAX12954	Enterococcus faeca
	32	75.5	16.4	3825	23	ABV25634	Human prostate exp
c	33	73	15.9	37856	21	AAA11992	S. cellulosum DNA
	34	71.5	15.6	1257	23	AAS80392	DNA encoding novel
	35	71.5	15.6	1257	23	AAS80800	DNA encoding novel
	36	71.5	15.6	1257	23	AAS81402	DNA encoding novel
c	37	71.5	15.6	2223	22	AAS00625	Human death-associ
	38	71.5	15.6	2271	23	AAS78101	DNA encoding novel
c	39	71.5	15.6	2340	19	AAV45136	Human Fas-binding
c	40	71.5	15.6	2477	22	AAF72902	Human daxx nucleot
	41	71.5	15.6	2682	21	AAA88863	Human dentin matri
	42	71.5	15.6	2811	23	AAS80952	DNA encoding novel
	43	71	15.5	3011	23	ABL07090	Drosophila melanog
	44	71	15.5	3341	23	ABL07130	Drosophila melanog
	45	71	15.5	32189	22	AAS30115	Human lung antigen

ALIGNMENTS

RESULT 1
AAT33677
ID AAT33677 standard; cDNA; 2406 BP.
XX
AC AAT33677;
XX
DT 18-NOV-1996 (first entry)
XX
DE Human transcription factor NFAT4a cDNA.
XX
KW Nuclear factor of activated T-cells; NFAT; NFAT4a;
KW transcription factor; binding assay; immune system disease;
KW gene therapy; diagnosis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 211..2337

```
FT                  /*tag= a
FT                  /product= NFAT4a
XX
PN   WO9626959-A1.
XX
PD   06-SEP-1996.
XX
PF   04-MAR-1996;   96WO-US03113.
XX
PR   02-MAR-1995;   95US-0396479.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Hoey T;
XX
DR   WPI; 1996-412738/41.
DR   P-PSDB; AAW02251.
XX
PT   DNA mol. encoding human nuclear factors of activated T cells -
PT   useful for screening potential therapeutic and diagnostic agents for
PT   immune system diseases
XX
PS   Disclosure; Page 50-52; 64pp; English.
XX
CC   A cDNA clone (AAT33677) codes for human nuclear factor of activated
CC   T-cells class 4a, NFAT4a (AAW02251), a regulator of immune system
CC   function.  2 Other classes of NFAT4 cDNAs, NFAT4b and NFAT4c,
CC   were also isolated (see also AAT36975 and AAT36871) that differed from
CC   NFAT4a at their 3' terminal sequences; the NFAT4c transcript was
CC   most abundunt.  The major sites for NFAT4 expression are in skeletal
CC   muscle and thymus.  NFAT cDNAs (see also AAT36866-68) were isolated
CC   from cDNA libraries prepd. from Jurkat T-cells and human peripheral
CC   blood lymphocytes by screening with a probe derived from the rel
CC   similarity region of the murine NFATp gene.  NFAT genes can be used
CC   for prodn. of recombinant NFATs, as probes and primers, and (esp.
CC   antisense) to modulate cellular expression or intracellular
CC   availability of active NFAT.
XX
SQ   Sequence 2406 BP; 645 A; 612 C; 535 G; 614 T; 0 other;

Alignment Scores:
Pred. No.:              2.13e-46        Length:         2406
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     17              Gaps:           0

US-09-550-115A-2 (1-86) x AAT33677 (1-2406)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||||:::|||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 2
AAF21183
ID   AAF21183 standard; DNA; 3583 BP.
XX
AC   AAF21183;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2750.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1103-1104; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 3583 BP; 959 A; 898 C; 757 G; 969 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              3.65e-46        Length:         3583
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     21              Gaps:           0

US-09-550-115A-2 (1-86) x AAF21183 (1-3583)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::||||||||:::|||||||||||||:::||||||||||||||||||||
Db   1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1225 TGGAGCAAACCAAAGCCT 1242

RESULT 3
AAA35061
ID   AAA35061 standard; DNA; 3583 BP.
XX
AC   AAA35061;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2750.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1025; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 3583 BP; 959 A; 898 C; 757 G; 969 T; 0 other;

Alignment Scores:
Pred. No.:              3.65e-46        Length:         3583
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     21              Gaps:           0

US-09-550-115A-2 (1-86) x AAA35061 (1-3583)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::||||||||:::|||||||||||||:::||||||||||||||||||||
Db   1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1225 TGGAGCAAACCAAAGCCT 1242

RESULT 4
AAF21182
ID   AAF21182 standard; DNA; 3769 BP.
XX
AC   AAF21182;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2749.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
```

KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1102-1103; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 3769 BP; 974 A; 939 C; 802 G; 1054 T; 0 other;

Alignment Scores:

Pred. No.:	3.91e-46	Length:	3769
Score:	430.00	Matches:	79
Percent Similarity:	96.51%	Conservative:	4
Best Local Similarity:	91.86%	Mismatches:	3
Query Match:	93.68%	Indels:	0
DB:	21	Gaps:	0

US-09-550-115A-2 (1-86) x AAF21182 (1-3769)

```
Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242
```

RESULT 5
AAA35060
ID AAA35060 standard; DNA; 3769 BP.
XX
AC AAA35060;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2749.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1024; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of

```
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 3769 BP; 974 A; 939 C; 802 G; 1054 T; 0 other;

Alignment Scores:
Pred. No.:                3.91e-46      Length:          3769
Score:                    430.00        Matches:         79
Percent Similarity:       96.51%        Conservative:    4
Best Local Similarity:    91.86%        Mismatches:      3
Query Match:              93.68%        Indels:          0
DB:                       21            Gaps:            0

US-09-550-115A-2 (1-86) x AAA35060 (1-3769)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||:::||||||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242

RESULT 6
AAF21181
ID   AAF21181 standard; DNA; 3784 BP.
XX
AC   AAF21181;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2748.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1101-1102; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 3784 BP; 979 A; 950 C; 804 G; 1051 T; 0 other;

Alignment Scores:
Pred. No.:                3.94e-46      Length:          3784
Score:                    430.00        Matches:         79
Percent Similarity:       96.51%        Conservative:    4
Best Local Similarity:    91.86%        Mismatches:      3
Query Match:              93.68%        Indels:          0
DB:                       21            Gaps:            0

US-09-550-115A-2 (1-86) x AAF21181 (1-3784)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||:::||||||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224
```

```
Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242

RESULT 7
AAA35059
ID   AAA35059 standard; DNA; 3784 BP.
XX
AC   AAA35059;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2748.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1023; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 3784 BP; 979 A; 950 C; 804 G; 1051 T; 0 other;

Alignment Scores:
Pred. No.:                3.94e-46      Length:          3784
Score:                    430.00        Matches:         79
Percent Similarity:       96.51%        Conservative:    4
Best Local Similarity:    91.86%        Mismatches:      3
Query Match:              93.68%        Indels:          0
DB:                       21            Gaps:            0

US-09-550-115A-2 (1-86) x AAA35059 (1-3784)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     985 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1044

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1045 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1104

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1105 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1164

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1165 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1224

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1225 TGGAGCAAACCAAAGCCT 1242

RESULT 8
ABN95715
ID   ABN95715 standard; DNA; 3969 BP.
XX
AC   ABN95715;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2213 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample -
XX
PS   Claim 1; SEQ ID NO 2213; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
```

```
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3969 BP; 993 A; 1021 C; 875 G; 1080 T; 0 other;

Alignment Scores:
Pred. No.:              4.2e-46       Length:          3969
Score:                  430.00        Matches:         79
Percent Similarity:     96.51%        Conservative:    4
Best Local Similarity:  91.86%        Mismatches:      3
Query Match:            93.68%        Indels:          0
DB:                     24            Gaps:            0

US-09-550-115A-2 (1-86) x ABN95715 (1-3969)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 9
AAQ84303
ID   AAQ84303 standard; cDNA; 4010 BP.
XX
AC   AAQ84303;
XX
DT   12-AUG-1995  (first entry)
XX
DE   Human NF-AT120, X subfamily gene.
XX
KW   NF-AT120; nuclear factor of activated T cells 120 protein;
KW   transcription; differentiation; cytokine; T-cell; T-lymphocyte; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             304..3531
FT                   /*tag= a
FT   unsure          1756..1758
FT                   /*tag= b
FT                   /note= "bases 1756-1758 may be replaced by gta,
FT                         gaa or gga"
FT   unsure          3090..3092
FT                   /*tag= c
FT                   /note= "bases 3090-3092 may be replaced by agt,
FT                         aga or agg"
XX
PN   WO9502053-A.
XX
PD   19-JAN-1995.
XX
PF   05-JUL-1994;  94WO-US07297.
XX
PR   06-JUL-1993;  93US-0088483.
PR   30-JUL-1993;  93US-0099998.
PR   30-AUG-1993;  93US-0113971.
PR   05-NOV-1993;  93US-0148061.
PR   04-APR-1994;  94US-0222626.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Arai N,  Masuda ES,  Tokumitsu H;
XX
DR   WPI; 1995-066898/09.
DR   P-PSDB; AAR75353.
XX
PT   Nuclear Factor of Activated T cells 120 protein, antibody and
PT   nucleic acid - useful as transcriptional regulator of genes, such
PT   as cytokines
XX
PS   Disclosure; Page 78-83; 100pp; English.
XX
CC   A DNA fragment amplified from Jurkat cDNA by PCR had the sequence
CC   given in AAQ84300 and encoded NF-AT120 (Nuclear Factor of
CC   Activated T cells protein 120) (AAR66873).  Amplified DNA was used
CC   to screen for related NF-AT proteins by hybridization, leading to
CC   the discovery of 3 related subfamilies, designated class C, P and
CC   X (AAR66874, AAR66875, AAR75353).
XX
SQ   Sequence 4010 BP; 1026 A; 1013 C; 902 G; 1069 T; 0 other;

Alignment Scores:
Pred. No.:              4.26e-46      Length:          4010
Score:                  430.00        Matches:         79
Percent Similarity:     96.51%        Conservative:    4
Best Local Similarity:  91.86%        Mismatches:      3
Query Match:            93.68%        Indels:          0
DB:                     16            Gaps:            0

US-09-550-115A-2 (1-86) x AAQ84303 (1-4010)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1264 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1323

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1324 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1383

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1384 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1443

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1444 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1503

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1504 TGGAGCAAACCAAAGCCT 1521

RESULT 10
AAF21185
ID   AAF21185 standard; DNA; 17069 BP.
XX
AC   AAF21185;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2752.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
```

```
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1105-1109; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oiigonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors; CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 17069 BP; 4137 A; 4938 C; 3964 G; 4030 T; 0 other;

Alignment Scores:
Pred. No.:              3.04e-45        Length:         17069
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     21              Gaps:           0

US-09-550-115A-2 (1-86) x AAF21185 (1-17069)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    3983 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 4042

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    4043 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 4102

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::||||||||:::||||||||||||:::||||||||||||||||||||
Db    4103 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 4162

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    4163 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 4222

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    4223 TGGAGCAAACCAAAGCCT 4240

RESULT 11
AAA35063
ID   AAA35063 standard; DNA; 17069 BP.
XX
AC   AAA35063;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2752.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1026-1031; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
```

```
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 17069 BP; 4137 A; 4938 C; 3964 G; 4030 T; 0 other;

Alignment Scores:
Pred. No.:              3.04e-45        Length:         17069
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     21              Gaps:           0

US-09-550-115A-2 (1-86) x AAA35063 (1-17069)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    3983 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 4042

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    4043 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 4102

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    4103 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 4162

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    4163 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 4222

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    4223 TGGAGCAAACCAAAGCCT 4240

RESULT 12
AAC14360/c
ID   AAC14360 standard; cDNA; 240 BP.
XX
AC   AAC14360;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 18435.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 18435; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 240 BP; 81 A; 49 C; 45 G; 59 T; 6 other;

Alignment Scores:
Pred. No.:              9.79e-16        Length:         240
Score:                  187.00          Matches:        34
Percent Similarity:     92.11%          Conservative:   1
Best Local Similarity:  89.47%          Mismatches:     3
Query Match:            40.74%          Indels:         0
DB:                     21              Gaps:           0

US-09-550-115A-2 (1-86) x AAC14360 (1-240)

Qy      49 ArgGluThrSerValAspAspGlyLeuGlySerGlnTyrProLeuLysLysAspSerSer 68
           ||||||||||||:::|||||||||||||||||||   ||||||||||||||||||||||
Db    234 CGGGAGACTTCAATAGATGATGGCCTTGGATCTCAKTATCCTTTAAAGAAAGATTCATGT 175

Qy      69 GlyAspGlnPheLeuSerValProSerProPheThrTrpSerLysProLysPro 86
           |||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db    174 GGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACCTGGAGCAAACCWAWKCCT 121

RESULT 13
AAT36868
ID   AAT36868 standard; cDNA; 2881 BP.
XX
AC   AAT36868;
XX
DT   17-NOV-1996  (first entry)
XX
DE   Human transcription factor NFAT3 cDNA.
XX
KW   Nuclear factor of activated T-cells; NFAT; NFAT3;
KW   transcription factor; binding assay; immune system disease;
KW   gene therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             142..2850
FT                   /*tag= a
FT                   /product= NFAT3
XX
PN   WO9626959-A1.
XX
PD   06-SEP-1996.
XX
PF   04-MAR-1996;   96WO-US03113.
XX
PR   02-MAR-1995;   95US-0396479.
```

```
XX
PA   (TULA-) TULARIK INC.
XX
PI   Hoey T;
XX
DR   WPI; 1996-412738/41.
DR   P-PSDB; AAW02250.
XX
PT   DNA mol. encoding human nuclear factors of activated T cells -
PT   useful for screening potential therapeutic and diagnostic agents for
PT   immune system diseases
XX
PS   Disclosure; Page 43-47; 64pp; English.
XX
CC   3 Overlapping cDNA clones were used to deduce a sequence (AAT36868)
CC   coding for human nuclear factor of activated T-cells class 3, NAFT3
CC   (AAW02250), a regulator of immune system function.  The major sites
CC   for NFAT3 expression are outside the immune system, with high
CC   expression in placenta, lung, kidney testis and ovary.  NFAT cDNAs
CC   (see also AAT36866-67, AAT33677, AAT36975 and AAT36871) were isolated
CC   from cDNA libraries prepd. from Jurkat T-cells and human peripheral
CC   blood lymphocytes by screening with a probe derived from the rel
CC   similarity region of the murine NFATp gene.  NFAT genes can be used
CC   for prodn. of recombinant NFATs, as probes and primers, and (esp.
CC   antisense) to modulate cellular expression or intracellular
CC   availability of active NFAT.
XX
SQ   Sequence 2881 BP; 531 A; 953 C; 882 G; 515 T; 0 other;

Alignment Scores:
Pred. No.:              1.05e-05      Length:         2881
Score:                  122.00        Matches:        34
Percent Similarity:     52.94%        Conservative:   11
Best Local Similarity:  40.00%        Mismatches:     34
Query Match:            26.58%        Indels:         6
DB:                     17            Gaps:           3

US-09-550-115A-2 (1-86) x AAT36868 (1-2881)

Qy      2 SerProAlaProPheProPheGlnTyrCysValGluThrAsp---IleProLeuLysThr 20
          ||||||   ||||||   :::                 :::   ||||||   ||||||
Db   1072 TCCCCTGGTCCCTTTGACTATGTGGGGGCCCCACCAGCTGAGAGCATCCCTCAGAAGACA 1131

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          |||:::||||||   :::||||||      ||||||         |||      |||
Db   1132 CGGCGGACTTCCAGCGAGCAGGCAGTGGCTCTGCCTCGGTCTGAGGAGCCTGCCTCATGC 1191

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          :::   |||   |||            |||   ||||||            ||||||
Db   1192 AAT---GGGAAGCTGCCCTTGGGAGCAGAGGAGTCTGTGGCTCCTCCAGGAGGTTCC--- 1245

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
                 :::|||:::   :::|||     :::|||:::||||||||||||
Db   1246 ---------CGGAAGGAGGTGGCTGGCATGGACTACCTGGCAGTGCCCTCCCCACTCGCT 1296

Qy     81 TrpSerLysProLys 85
          |||||||||   :::
Db   1297 TGGTCCAAGGCCCGG 1311

RESULT 14
AAX08717
ID   AAX08717 standard; DNA; 2881 BP.
XX
AC   AAX08717;
XX
DT   27-SEP-1999  (first entry)
XX
DE   NF-AT3 encoding sequence.
XX
KW   NF-AT3; hypertrophy; cardiomyocytes; cardiac hypertrophic response;
KW   heart failure; transgenic animals; screening; treatment; inhibition;
KW   ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             142..2850
FT                   /*tag=  a
FT                   /product= "NF-AT3"
FT   misc_difference 181..183
FT                   /*tag=  b
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 1126..1128
FT                   /*tag=  c
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 1519..1521
FT                   /*tag=  d
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 1609..1611
FT                   /*tag=  e
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 1654..1656
FT                   /*tag=  f
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 1723..1725
FT                   /*tag=  g
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 1987..1989
FT                   /*tag=  h
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 2020..2022
FT                   /*tag=  i
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 2107..2109
FT                   /*tag=  j
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 2207..2209
FT                   /*tag=  k
FT                   /transl_except= AAG encodes Leu
FT   misc_difference 2665..2667
FT                   /*tag=  l
FT                   /transl_except= AAG encodes Leu
XX
PN   WO9919471-A1.
XX
PD   22-APR-1999.
XX
PF   15-OCT-1998;   98WO-US21845.
XX
PR   16-APR-1998;   98US-0061417.
PR   16-OCT-1997;   97US-0062864.
PR   10-NOV-1997;   97US-0065178.
PR   15-APR-1998;   98US-0081853.
XX
PA   (UYNT-) UNIV NORTH TEXAS HEALTH SCI CENT.
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Grant SR,  Molkentin JD,  Olson EN;
XX
DR   WPI; 1999-277635/23.
DR   P-PSDB; AAW85730.
XX
PT   Treating hypertrophy in cardiomyocytes by inhibiting NF-A3
XX
PS   Disclosure; Page 99-101; 105pp; English.
XX
CC   Hypertrophy in cardiomyocytes is treated by inhibiting function of
CC   NF-AT3.  Activation of NF-AT3 mediates the calcium ion-dependent
CC   cardiac hypertrophic response to a variety of stimuli, so inhibiting
CC   it can be used to treat or prevent cardiac hypertrophy and related
CC   heart failure. Transgenic animals, or cells, containing a
CC   constitutively active NF-AT3 gene can be used as models for screening
CC   modulators of hypertrophy and for studying human disease.
XX
SQ   Sequence 2881 BP; 531 A; 953 C; 882 G; 515 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              1.05e-05        Length:         2881
Score:                  122.00          Matches:        34
Percent Similarity:     52.94%          Conservative:   11
Best Local Similarity:  40.00%          Mismatches:     34
Query Match:            26.58%          Indels:         6
DB:                     20              Gaps:           3

US-09-550-115A-2 (1-86) x AAX08717 (1-2881)

Qy      2 SerProAlaProPheProPheGlnTyrCysValGluThrAsp---IleProLeuLysThr 20
          ||||||   ||||||   :::                  :::   ||||||   ||||||
Db   1072 TCCCCTGGTCCCTTTGACTATGTGGGGGCCCCACCAGCTGAGAGCATCCCTCAGAAGACA 1131

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          |||:::||||||   :::||||||      ||||||         |||      |||
Db   1132 CGGCGGACTTCCAGCGAGCAGGCAGTGGCTCTGCCTCGGTCTGAGGAGCCTGCCTCATGC 1191

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          :::   |||   |||           |||   ||||||            ||||||
Db   1192 AAT---GGGAAGCTGCCCTTGGGAGCAGAGGAGTCTGTGGCTCCTCCAGGAGGTTCC--- 1245

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
                :::|||:::   :::|||      :::|||:::|||||||||||
Db   1246 ---------CGGAAGGAGGTGGCTGGCATGGACTACCTGGCAGTGCCCTCCCCACTCGCT 1296

Qy     81 TrpSerLysProLys 85
          |||||||||   :::
Db   1297 TGGTCCAAGGCCCGG 1311

RESULT 15
AAT36867
ID   AAT36867 standard; cDNA; 2743 BP.
XX
AC   AAT36867;
XX
DT   17-NOV-1996  (first entry)
XX
DE   Human transcription factor NFATc cDNA.
XX
KW   Nuclear factor of activated T-cells; NFAT; NFATc;
KW   transcription factor; binding assay; immune system disease;
KW   gene therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             240..2390
FT                   /*tag= a
FT                   /product= NFATc
XX
PN   WO9626959-A1.
XX
PD   06-SEP-1996.
XX
PF   04-MAR-1996;   96WO-US03113.
XX
PR   02-MAR-1995;   95US-0396479.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Hoey T;
XX
DR   WPI; 1996-412738/41.
DR   P-PSDB; AAW02249.
XX
PT   DNA mol. encoding human nuclear factors of activated T cells -
PT   useful for screening potential therapeutic and diagnostic agents for
PT   immune system diseases
XX
PS   Disclosure; Page 38-41; 64pp; English.
XX
CC   A cDNA clone (AAT36867) codes for human nuclear factor of activated
CC   T-cells class c, NAFTc (AAW02249), a regulator of immune system
CC   function.  NAFTc was expressed at a low level in several human
CC   tissues; a 4.5 kb transcript was strongly expressed in skeletal
CC   muscle.  NFAT cDNAs (see also AAT36866, AAT36868, AAT33677, AAT36975 and
CC   AAT36871) were isolated from cDNA libraries prepd. from Jurkat T-cells
CC   and human peripheral blood lymphocytes by screening with a probe
CC   derived from the rel similarity region of the murine NFATp gene.
CC   NFAT genes can be used for prodn. of recombinant NFATs, as probes
CC   and primers, and (esp. antisense) to modulate cellular expression
CC   or intracellular availability of active NFAT.
XX
SQ   Sequence 2743 BP; 529 A; 959 C; 788 G; 467 T; 0 other;

Alignment Scores:
Pred. No.:              0.000111        Length:         2743
Score:                  114.00          Matches:        25
Percent Similarity:     61.11%          Conservative:   19
Best Local Similarity:  34.72%          Mismatches:     26
Query Match:            24.84%          Indels:         2
DB:                     17              Gaps:           2

US-09-550-115A-2 (1-86) x AAT36867 (1-2743)

Qy     16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
          :::|||:::|||:::||||||||||:::   :::|||        :::      |||:::
Db   1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCCGCCCTCAGTGGCGCTCAAGGTG 1289

Qy     36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
          |||         :::|||   |||:::      |||      :::   :::   :::|||
Db   1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy     56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                  |||:::   :::::|||           ||||||:::|||:::|||
Db   1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy     76 ProSer---ProPheThrTrpSerLysProLysPro 86
          |||      |||:::   |||:::||||||||||||
Db   1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442


Search completed: April 18, 2003, 23:04:27
Job time : 252 secs
```

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 18, 2003, 22:06:52 ; Search time 1723 Seconds
(without alignments)
808.364 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09550115/runat_11042003_141903_18158/app_query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09550115_@CGN_1_1_1906_@runat_11042003_141903_18158 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	459	100.0	600	12	BG805976	BG805976 2012-19 M
2	459	100.0	600	13	BI990851	BI990851 4082-17 M
c 3	449	97.8	1073	9	AA821465	AA821465 vs25h04.r
c 4	430	93.7	455	14	BQ307368	BQ307368 MR0-BT400
c 5	430	93.7	483	14	BQ307397	BQ307397 MR0-BT400
6	430	93.7	1002	14	BQ070507	BQ070507 AGENCOURT
c 7	422	91.9	484	14	BQ307403	BQ307403 MR0-BT400
8	418	91.1	502	14	BQ307520	BQ307520 MR0-BT400
9	417	90.8	669	9	AU129466	AU129466 AU129466
10	413	90.0	483	14	BQ307401	BQ307401 MR0-BT400
11	411	89.5	486	14	BQ307399	BQ307399 MR0-BT400
12	410	89.3	920	14	BQ423593	BQ423593 AGENCOURT
c 13	409	89.1	497	14	BQ307525	BQ307525 MR0-BT400
c 14	409	89.1	808	13	BI408176	BI408176 602918865
c 15	402	87.6	379	10	AW244381	AW244381 BR_END03D
16	359	78.2	924	14	BQ214451	BQ214451 AGENCOURT
c 17	353	76.9	505	14	BQ307486	BQ307486 MR0-BT400
18	328	71.5	501	9	AA840427	AA840427 vx93a04.r
c 19	316	68.8	1067	13	BI525206	BI525206 602924609
20	313	68.2	728	9	AU123215	AU123215 AU123215
21	259	56.4	490	9	AL120778	AL120778 DKFZp762C
c 22	247	53.8	435	12	BF566087	BF566087 UI-R-BT1-
c 23	209	45.5	918	17	CNS05BC3	AL329628 Tetraodon
24	164.5	35.8	369	9	AA188523	AA188523 zp74c08.r
25	154	33.6	264	13	BJ095964	BJ095964 BJ095964
c 26	145.5	31.7	573	9	AI016619	AI016619 ov30e08.x
27	129	28.1	723	13	BI112028	BI112028 602901368
28	129	28.1	844	11	AK014164	AK014164 Mus muscu
29	122	26.6	1482	13	BM558073	BM558073 AGENCOURT
c 30	117	25.5	259	9	AA770124	AA770124 ah82c05.s
31	114	24.8	1024	14	BQ710263	BQ710263 AGENCOURT
32	107	23.3	1044	14	BQ714666	BQ714666 AGENCOURT
33	105	22.9	449	12	BF776523	BF776523 287350 MA
34	105	22.9	513	13	BM105571	BM105571 509057 MA
35	103	22.4	461	12	BG746287	BG746287 602703517
36	96.5	21.0	4077	11	AK004810	AK004810 Mus muscu
37	93	20.3	530	14	R87618	R87618 ym91g05.r1
38	83	18.1	437	9	AI750778	AI750778 cn05e04.x
39	82	17.9	877	14	BQ733684	BQ733684 AGENCOURT
40	81.5	17.8	594	10	AV827456	AV827456 AV827456
41	81.5	17.8	903	10	BE616886	BE616886 601278756
42	80	17.4	897	9	AL522401	AL522401 AL522401
c 43	78.5	17.1	1074	17	CNS03LNG	AL249685 Tetraodon
c 44	78	17.0	1169	12	BE899396	BE899396 601681370
c 45	77	16.8	653	13	BM668476	BM668476 UI-E-CL1-

ALIGNMENTS

RESULT 1
BG805976
LOCUS BG805976 600 bp mRNA linear EST 20-DEC-2001
DEFINITION 2012-19 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.
ACCESSION BG805976
VERSION BG805976.1 GI:17952916
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W., White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.

```
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
                     Primer: Oligo dT. RNA Isolation: cytoplasmic  RNA preps
                     (Manniatis); Cloning Technique: CUA Cloning (CloneAmp,
                     Life  Technologies); Average insert size: 1.8 Kb;
                     Insertion site:  TACGTCCACTGAATTCTGAGTG--->.  Other
                     information regarding entire library  may be found at
                     http://pga.swmed.edu/Data/Libraries/microarray_cdna_librar
                     ies.htm."
BASE COUNT      171 a    150 c    121 g    158 t
ORIGIN

Alignment Scores:
Pred. No.:              1.15e-47       Length:         600
Score:                  459.00         Matches:        86
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     12             Gaps:           0

US-09-550-115A-2 (1-86) x BG805976 (1-600)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 94

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95 AGGAAGACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGATCTGTTCAGAT 154

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155 GATCAAGGGAACTTATCCCCTTCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 214

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCCTTTACC 274

Qy      81 TrpSerLysProLysPro 86
           ||||||||||||||||||
Db     275 TGGAGCAAACCAAAGCCT 292

RESULT 2
BI990851
LOCUS       BI990851                600 bp    mRNA    linear   EST 20-DEC-2001
DEFINITION  4082-17 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BI990851
VERSION     BI990851.1  GI:17961868
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
```

```
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
                     Primer: Oligo dT. RNA Isolation: cytoplasmic  RNA preps
                     (Manniatis); Cloning Technique: CUA Cloning (CloneAmp,
                     Life  Technologies); Average insert size: 1.8 Kb;
                     Insertion site:  TACGTCCACTGAATTCTGAGTG--->.  Other
                     information regarding entire library  may be found at
                     http://pga.swmed.edu/Data/Libraries/microarray_cdna_librar
                     ies.htm."
BASE COUNT      158 a    167 c    118 g    157 t
ORIGIN

Alignment Scores:
Pred. No.:              1.15e-47       Length:         600
Score:                  459.00         Matches:        86
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     13             Gaps:           0

US-09-550-115A-2 (1-86) x BI990851 (1-600)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 222

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 AGGAAGACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGATCTGTTCAGAT 282

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 GATCAAGGGAACTTATCCCCTTCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 342

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCCTTTACC 402

Qy      81 TrpSerLysProLysPro 86
           ||||||||||||||||||
Db     403 TGGAGCAAACCAAAGCCT 420

RESULT 3
AA821465/c
LOCUS       AA821465                1073 bp    mRNA    linear   EST 17-FEB-1998
DEFINITION  vs25h04.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1139287 5' similar to gb:U28807 Mus musculus
            lymphoid-specific transcription factor NFATc3 (MOUSE);, mRNA
            sequence.
ACCESSION   AA821465
VERSION     AA821465.1  GI:2891333
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1073)
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:620559
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .1073
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1139287"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. The C2C12 cell line
                     (available from ATCC, catalog # CRL-1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins. "
BASE COUNT      299 a    213 c    291 g    270 t
ORIGIN

Alignment Scores:
Pred. No.:              4.75e-46        Length:         1073
Score:                  449.00          Matches:        84
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  98.82%          Mismatches:     0
Query Match:            97.82%          Indels:         0
DB:                     9               Gaps:           0

US-09-550-115A-2 (1-86) x AA821465 (1-1073)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     262 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 203

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     202 AGGAAGACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGATCTGTTCAGAT 143

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 GATCAAGGGAACTTATCCCCTTCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 83

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCCTTTACC 23

Qy      81 TrpSerLysProLys 85
           |||||||||||:::
Db      22 TGGAGCAAACCAAGG 8

RESULT 4
```

```
BQ307368/c
LOCUS       BQ307368                455 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  MR0-BT4000-190601-101-b08 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ307368
VERSION     BQ307368.1  GI:20843385
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT4000-
            190601-101-b08&t3=2001-06-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 8.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT4000"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      122 a     94 c    116 g    123 t
ORIGIN

Alignment Scores:
Pred. No.:              3.54e-44        Length:         455
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     14              Gaps:           0

US-09-550-115A-2 (1-86) x BQ307368 (1-455)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db     374 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAGACA 315

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     314 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 255

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db     254 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 195
```

```
Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db     194 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 135

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db     134 TGGAGCAAACCAAAGCCT 117

RESULT 5
BQ307397/c
LOCUS       BQ307397                483 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  MR0-BT4000-190601-101-a06 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ307397
VERSION     BQ307397.1  GI:20843470
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT4000-
            190601-101-a06&t3=2001-06-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 483.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT4000"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      130 a    102 c    121 g    130 t
ORIGIN

Alignment Scores:
Pred. No.:              3.85e-44        Length:         483
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     14              Gaps:           0

US-09-550-115A-2 (1-86) x BQ307397 (1-483)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db     402 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 343

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     342 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 283

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||:::||||||||:::|||||||||||:::||||||||||||||||||||||||
Db     282 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 223

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db     222 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 163

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db     162 TGGAGCAAACCAAAGCCT 145

RESULT 6
BQ070507
LOCUS       BQ070507                1002 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6839210 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5922897
            5', mRNA sequence.
ACCESSION   BQ070507
VERSION     BQ070507.1  GI:19899553
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1002)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2090  row: i  column: 10
            High quality sequence stop: 687.
FEATURES             Location/Qualifiers
     source          1. .1002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5922897"
                     /clone_lib="NIH_MGC_47"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      264 a    264 c    221 g    251 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              1.07e-43        Length:         1002
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     14              Gaps:           0

US-09-550-115A-2 (1-86) x BQ070507 (1-1002)
```

```
Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     379 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 438

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     439 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 498

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::||||||||||||||||||||||
Db     499 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 558

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db     559 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 618

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db     619 TGGAGCAAACCAAAGCCT 636
```

RESULT 7
BQ307403/c

LOCUS	BQ307403 484 bp mRNA linear EST 16-MAY-2002
DEFINITION	MR0-BT4000-190601-101-d06 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BQ307403
VERSION	BQ307403.1 GI:20843484
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 484)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT4000-190601-101-d06&t3=2001-06-19&t4=1) Seq primer: puc 18 forward High quality sequence start: 5 High quality sequence stop: 484.

FEATURES	Location/Qualifiers
source	1. .484 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BT4000" /dev_stage="Adult" /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 129 a 103 c 121 g 131 t
ORIGIN

Alignment Scores:

Pred. No.:	3.94e-43	Length:	484
Score:	422.00	Matches:	78
Percent Similarity:	95.35%	Conservative:	4
Best Local Similarity:	90.70%	Mismatches:	4
Query Match:	91.94%	Indels:	0
DB:	14	Gaps:	0

US-09-550-115A-2 (1-86) x BQ307403 (1-484)

```
Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db     403 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 344

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     343 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 284

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::||||||||||||||||||||||
Db     283 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 224

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   |||||||||   ||||||||||||||||||||||||||
Db     223 TATCCTTTAAAGAAGGATTCATGTGGTGATCAGTCTCTTTCAGTTCCTTCACCCTTTACC 164

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db     163 TGGAGCAAACCAAAGCCT 146
```

RESULT 8
BQ307520

LOCUS	BQ307520 502 bp mRNA linear EST 16-MAY-2002
DEFINITION	MR0-BT4000-190601-102-f11 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BQ307520
VERSION	BQ307520.1 GI:20843835
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 502)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT4000-190601-102-f11&t3=2001-06-19&t4=1) Seq primer: puc 18 forward High quality sequence start: 6 High quality sequence stop: 502.

FEATURES	Location/Qualifiers
source	1. .502 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BT4000"

```
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      137 a    133 c    101 g    131 t
ORIGIN

Alignment Scores:
Pred. No.:              1.33e-42        Length:         502
Score:                  418.00          Matches:        77
Percent Similarity:     96.43%          Conservative:   4
Best Local Similarity:  91.67%          Mismatches:     3
Query Match:            91.07%          Indels:         0
DB:                     14              Gaps:           0

US-09-550-115A-2 (1-86) x BQ307520 (1-502)

Qy       3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
           ||||||   ||||||   ||||||||||||||||||||||||||||||||||||||||||
Db     117 CCTGCAGTTTTTCCATCTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACAAGGAAA 176

Qy      23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAspGln 42
           ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db     177 ACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGATGACCAA 236

Qy      43 GlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGlnTyrPro 62
           |||:::|||||||||:::|||||||||||:::|||||||||||||||||||||||||||
Db     237 GGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAGTATCCT 296

Qy      63 LeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThrTrpSer 82
           ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db     297 TTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACCTGGAGC 356

Qy      83 LysProLysPro 86
           ||||||||||||
Db     357 AAACCAAAGCCT 368

RESULT 9
AU129466
LOCUS       AU129466                669 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU129466 NT2RP2 Homo sapiens cDNA clone NT2RP2005598 5', mRNA
            sequence.
ACCESSION   AU129466
VERSION     AU129466.1  GI:10989820
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2005598"
                     /clone_lib="NT2RP2"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT      202 a    155 c    121 g    179 t     12 others
ORIGIN

Alignment Scores:
Pred. No.:              2.66e-42        Length:         669
Score:                  417.00          Matches:        76
Percent Similarity:     98.77%          Conservative:   4
Best Local Similarity:  93.83%          Mismatches:     1
Query Match:            90.85%          Indels:         0
DB:                     9               Gaps:           0

US-09-550-115A-2 (1-86) x AU129466 (1-669)

Qy       6 PheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLysThrSerGlu 25
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 TTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACAAGGAAAACTTCTGAA 61

Qy      26 AspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeu 45
           ||||||||||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db      62 GATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGATGACCAAGGGAGTTTA 121

Qy      46 SerProSerArgGluThrSerValAspAspGlyLeuGlySerGlnTyrProLeuLysLys 65
           ||||||:::||||||||||||:::|||||||||||||||||||||||||||||||||||
Db     122 TCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAGTATCCTTTAAAGAAA 181

Qy      66 AspSerSerGlyAspGlnPheLeuSerValProSerProPheThrTrpSerLysProLys 85
           ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACCTGGAGCAAACCAAAG 241

Qy      86 Pro 86
           |||
Db     242 CCT 244

RESULT 10
BQ307401
LOCUS       BQ307401                483 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  MR0-BT4000-190601-101-c10 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ307401
VERSION     BQ307401.1  GI:20843478
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT4000-
```

```
           190601-101-c10&t3=2001-06-19&t4=1)
           Seq primer: puc 18 forward
           High quality sequence stop: 483.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT4000"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      134 a    126 c     97 g    125 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                5.37e-42       Length:          483
Score:                    413.00         Matches:         76
Percent Similarity:       95.29%         Conservative:    5
Best Local Similarity:    89.41%         Mismatches:      4
Query Match:              89.98%         Indels:          0
DB:                       14             Gaps:            0

US-09-550-115A-2 (1-86) x BQ307401 (1-483)

Qy      2 SerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArg 21
          :::|||||||   ||||||||||||||||||||||||   ||||||||||||||||||||
Db     98 GCGCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGGCTGACATCCCTCTCAAAACAAGG 157

Qy     22 LysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAsp 41
          |||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db    158 AAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGATGAC 217

Qy     42 GlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGlnTyr 61
          ||||||:::||||||||||:::||||||||||||:::|||||||||||||||||||||||
Db    218 CAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCACAGTAT 277

Qy     62 ProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThrTrp 81
          ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db    278 CCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACCTGG 337

Qy     82 SerLysProLysPro 86
          |||||||  ||||||
Db    338 AGCAAANCAAAGCCT 352

RESULT 11
BQ307399
LOCUS       BQ307399                 486 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  MR0-BT4000-190601-101-b10 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ307399
VERSION     BQ307399.1  GI:20843474
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-BT4000-
            190601-101-b10&t3=2001-06-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 486.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT4000"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      135 a    126 c    100 g    125 t
ORIGIN

Alignment Scores:
Pred. No.:                9.68e-42       Length:          486
Score:                    411.00         Matches:         77
Percent Similarity:       96.47%         Conservative:    5
Best Local Similarity:    90.59%         Mismatches:      2
Query Match:              89.54%         Indels:          1
DB:                       14             Gaps:            0

US-09-550-115A-2 (1-86) x BQ307399 (1-486)

Qy      3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 CCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACAAGGAAA 160

Qy     23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAspGln 42
          |||||||||||||||||||||||||||||||||||||||||:::|||:::||||||||||
Db    161 ACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTACAGATGACCAA 220

Qy     43 GlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln-TyrPr 62
          |||:::|||||||||:::||||||||||||:::|||||||||||||||||||| |||||
Db    221 GGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAGTTATCC 280

Qy     62 oLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThrTrpSe 82
          ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db    281 TTTAAAGAAAGATTCAGGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACCTGGAG 340

Qy     82 rLysProLysPro 86
          ||||||||||||
Db    341 CAAACCAAAGCCG 353

RESULT 12
BQ423593
LOCUS       BQ423593                 920 bp    mRNA    linear   EST 23-MAY-2002
DEFINITION  AGENCOURT_7762469 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6068527
            5', mRNA sequence.
ACCESSION   BQ423593
VERSION     BQ423593.1  GI:21118908
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13349  row: m  column: 08
            High quality sequence stop: 619.
FEATURES             Location/Qualifiers
     source          1. .920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6068527"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      222 a    250 c    208 g    240 t
ORIGIN

Alignment Scores:
Pred. No.:              3.17e-41        Length:         920
Score:                  410.00          Matches:        78
Percent Similarity:     94.25%          Conservative:   4
Best Local Similarity:  89.66%          Mismatches:     4
Query Match:            89.32%          Indels:         1
DB:                     14              Gaps:           0

US-09-550-115A-2 (1-86) x BQ423593 (1-920)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db    487 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 546

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    547 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 606

Qy     41 AspGln-GlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGl 60
          |||||| |||:::||||||||||:::|||||||||||||:::|||||||||||||||||||
Db    607 GACCAAGGGAAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCA 666

Qy     60 nTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheTh 80
          ||||||||||||||||||||   |||||||||   ||||||||||||||||||||||||
Db    667 GTATCCTTTAAAGAAAGATTCATGTGGTGATCAGGTTCTTTCAGTTCCTTCACCCTTTAC 726

Qy     80 rTrpSerLysProLysPro 86
          ||||||||||||||||||
Db    727 CTGGAGCAAACCAAAGCCT 745

RESULT 13
BQ307525/c
LOCUS       BQ307525                497 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  MR0-BT4000-190601-102-h04 BT4000 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ307525
VERSION     BQ307525.1  GI:20843850
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-BT4000-
            190601-102-h04&t3=2001-06-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT4000"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      132 a    108 c    123 g    133 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.79e-41        Length:         497
Score:                  409.00          Matches:        75
Percent Similarity:     96.43%          Conservative:   6
Best Local Similarity:  89.29%          Mismatches:     3
Query Match:            89.11%          Indels:         0
DB:                     14              Gaps:           0

US-09-550-115A-2 (1-86) x BQ307525 (1-497)

Qy      3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
          ||||||   ||||||||||||||||||||||||||||||||||||:::|||||||||
Db    407 CCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCGAAACAAGGAAA 348

Qy     23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAspGln 42
          ||||||||||||||||||||||||||||||||||||||:::|||||||||||||:::
Db    347 ACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGATGACCGA 288

Qy     43 GlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGlnTyrPro 62
          |||:::||||||   :::|||||||||||:::|||||||||||||||||||||||||||
Db    287 GGGAGTTTATCANCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAGTATCCT 228

Qy     63 LeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThrTrpSer 82
          ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db    227 TTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACCTGGAGC 168

Qy     83 LysProLysPro 86
          ||||||||||||
Db    167 AAACCAAAGCCT 156

RESULT 14
BI408176/c
LOCUS       BI408176                808 bp    mRNA    linear   EST 14-AUG-2001
DEFINITION  602918865F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5055641 5',
            mRNA sequence.
```

```
ACCESSION   BI408176
VERSION     BI408176.1  GI:15169099
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11151  row: a  column: 18
            High quality sequence start: 26
            High quality sequence stop: 796.
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5055641"
                     /clone_lib="NCI_CGAP_Lu33"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia),  digested with Not I and cloned into the Not
                     I and Eco RI sites of  the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      226 a    167 c    229 g    186 t
ORIGIN

Alignment Scores:
Pred. No.:                3.54e-41        Length:          808
Score:                    409.00          Matches:         78
Percent Similarity:       98.73%          Conservative:    0
Best Local Similarity:    98.73%          Mismatches:      1
Query Match:              89.11%          Indels:          0
DB:                       13              Gaps:            0

US-09-550-115A-2 (1-86) x BI408176 (1-808)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     260 CTCAGCCCTGCACCGTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACA 201

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 AGGAAGACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGATCTGTTCAGAT 141

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     140 GATCAAGGGAACTTATCCCCTTCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAG 81

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPhe 79
           ||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db      80 TATCCTTTAAAGAAAGATTCATCTGGTGACCAATTTCGTTCAGTTCCTTCACCCTTT 24

RESULT 15
AW244381/c
LOCUS       AW244381                379 bp    mRNA    linear   EST 25-JAN-2002
DEFINITION  BR_END03D12 Bain Rancourt retinoic acid induced ES cell neural
            differentiation subtraction library Mus musculus cDNA clone 03D12
            similar to dbj|D85612|D85612 Mouse mRNA for NFATx, complete cds and
            gb|U28807|MMU28807 Mus musculus lymphoidspecific transcription
            factor NFATc3 mRNA, partial cds, mRNA sequence.
ACCESSION   AW244381
VERSION     AW244381.1  GI:8051133
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Bain,G., Mansergh,F.C., Wride,M.A., Hance,J.E., Isogawa,A.,
            Rancourt,S.L., Ray,W.J., Yoshimura,Y., Tsuzuki,T., Gottlieb,D.I.
            and Rancourt,D.E.
  TITLE     ES cell neural differentiation reveals a substantial number of
            novel ESTs
  JOURNAL   Funct. Integr. Genomics 1 (2), 127-139 (2000)
  MEDLINE   21652683
COMMENT     Contact: Rancourt DE
            Department of Biochemistry and Molecular Biology
            University of Calgary
            3330 Hospital Drive N.W., Calgary, Alberta, T2N 4N1, Canada
            Tel: 403 220 2888
            Fax: 403 283 8727
            Email: rancourt@ucalgary.ca; URL: http://www.acs.ucalgary.ca/(tilde
            )rancourt/
            DNA sequencing by: University Core DNA Services, University of
            Calgary. Submitted sequence has been trimmed at both ends to remove
            the adaptor oligos containing the EcoRI sites; i.e. GAATTCGGACTA
            (beginning) and TAGTCCGAATTC (end) removed. Therefore, reported
            insert length is longer than actual EST sequence length.
            Insert Length: 389   Std Error: 10.00
            Seq primer: T3 Or T7.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Mus musculus"
                     /strain="129"
                     /db_xref="taxon:10090"
                     /clone="03D12"
                     /clone_lib="Bain Rancourt retinoic acid induced ES cell
                     neural differentiation subtraction library"
                     /cell_type="embryonic stem (ES) cell"
                     /cell_line="D3"
                     /lab_host="DH5 alpha"
                     /note="Vector: pBluescript II SK+ (Stratagene); Site_1:
                     EcoR1; Site_2: EcoR1; Library constructed by Dr. Gerard
                     Bain (present address: Hoechst-ARIAD  Genomic Center,
                     ARIAD Pharmaceuticals Inc., 26 Landsdowne Street,
                     Cambridge, Massachusetts, 02139-4234, U.S.A.). To isolate
                     cDNAs  corresponding to mRNAs which are upregulated during
                     the neural  differentiation of ES cells in vitro, the
                     subtractive hybridization  technique of Wang and Brown [1]
                     was employed. Poly(A)+ RNA was  prepared from both
                     undifferentiated ES cells and from embryoid bodies  which
                     had been cultured for 4 days in the absence of RA followed
                     by an additional 3 days in the presence of 0.5 (M RA
                     (4-/3+ cells).  These poly(A)+ RNAs were converted to
                     double-stranded cDNA using the  Superscript Choice System
                     (Gibco). Aliquots of both cDNAs were  digested with the
                     restriction enzymes AluI and AluI plus RsaI. An  adaptor
                     oligo [1] containing an EcoRI site was ligated to the ends
                     of  the restricted cDNAs to provide primer binding sites
                     and large amounts of  each cDNA population were then
                     produced by the polymerase chain  reaction (PCR) as
                     described [1]. Amplified cDNA from undifferentiated  ES
                     cells was biotinylated using Photoprobe biotin (Vector
                     Laboratories)  according to the manufacturer's protocol.
                     2.5 ug of amplified cDNA  from 4-/3+ cells was mixed with
                     50 ug of biotinylated ES cell cDNA,  denatured by boiling,
```

and hybridized for 20 h. Double stranded cDNAs containing biotin were removed by streptavidin/phenol treatment as described [1]. The remaining subtracted cDNA was mixed with an additional 25 mg of biotinylated ES cell cDNA, denatured by boiling, and hybridized for 2 h. The streptavidin/phenol treatment was repeated and the remaining cDNA was amplified by PCR [Wang and Brown, 1991]. Two additional rounds of subtraction were repeated exactly as described above. The cDNA obtained from this subtraction procedure was digested with EcoRI and ligated to pBS II SK+ (Stratagene) followed by transformation into E. coli DH5 cells. Individual colonies were picked and the corresponding plasmids were isolated either by an alkaline lysis miniprep procedure [2], or using the Qiaprep spin miniprep kit (Qiagen). Sequence analysis was performed using the Big Dye Cycle Sequencing kit and an ABI373 sequencer (University Core DNA Services, University of Calgary). 1. Wang,Z; Brown,DD (1991) A gene expression screen. Proc. Natl. Acad. Sc i. USA 88, 11505-11509. 2. Sambrook,J; Fritsch,EF; Maniatis,T. (1989) Molecular Cloning:A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor,New York."

BASE COUNT 124 a 65 c 95 g 95 t
ORIGIN

Alignment Scores:

Pred. No.:	9.31e-41	Length:	379
Score:	402.00	Matches:	76
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	87.58%	Indels:	0
DB:	10	Gaps:	0

US-09-550-115A-2 (1-86) x AW244381 (1-379)

```
Qy      11 CysValGluThrAspIleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIle 30
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 TGTGTAGAGACTGACATCCCTTTGAAAACAAGGAAGACTTCTGAAGATCAAGCTGCCATA 317

Qy      31 LeuProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGlu 50
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     316 CTACCAGGAAAATTAGAGATCTGTTCAGATGATCAAGGGAACTTATCCCCTTCCCGGGAG 257

Qy      51 ThrSerValAspAspGlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAsp 70
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACATCAGTAGATGATGGCCTTGGATCTCAGTATCCTTTAAAGAAAGATTCATCTGGTGAC 197

Qy      71 GlnPheLeuSerValProSerProPheThrTrpSerLysProLysPro 86
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 CAATTTCTTTCAGTTCCTTCACCCTTTACCTGGAGCAAACCAAAGCCT 149
```

Search completed: April 18, 2003, 23:53:27
Job time : 1734 secs

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 18, 2003, 22:21:47 ; Search time 53 Seconds
(without alignments)
497.626 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09550115/runat_11042003_141904_18174/app_query.fasta_1.263
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09550115_@CGN_1_1_40_@runat_11042003_141904_18174 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	430	93.7	2406	1	US-08-396-479B-7	Sequence 7, Appli
2	430	93.7	2406	1	US-08-818-823-7	Sequence 7, Appli
3	430	93.7	2647	1	US-08-396-479B-9	Sequence 9, Appli
4	430	93.7	2647	1	US-08-818-823-9	Sequence 9, Appli
5	430	93.7	3969	1	US-08-396-479B-11	Sequence 11, Appl
6	430	93.7	3969	1	US-08-818-823-11	Sequence 11, Appl
7	430	93.7	4010	5	PCT-US94-07297-40	Sequence 40, Appl
8	122	26.6	2881	1	US-08-396-479B-5	Sequence 5, Appli
9	122	26.6	2881	1	US-08-818-823-5	Sequence 5, Appli
10	114	24.8	2743	1	US-08-396-479B-3	Sequence 3, Appli
11	114	24.8	2743	1	US-08-818-823-3	Sequence 3, Appli
12	114	24.8	2749	2	US-08-124-981A-1	Sequence 1, Appli
13	114	24.8	2751	3	US-09-037-190-45	Sequence 45, Appl
14	114	24.8	2751	3	US-09-037-192-45	Sequence 45, Appl
15	114	24.8	2751	3	US-09-037-143-45	Sequence 45, Appl
16	114	24.8	2751	4	US-09-049-691-45	Sequence 45, Appl
17	114	24.8	2751	4	US-08-260-174-45	Sequence 45, Appl
18	114	24.8	2751	4	US-09-338-128A-45	Sequence 45, Appl
19	114	24.8	2751	4	US-09-037-192-45	Sequence 45, Appl
20	114	24.8	2853	5	PCT-US94-07297-36	Sequence 36, Appl
21	114	24.8	5502	4	US-09-232-346-45	Sequence 45, Appl
c 22	71.5	15.6	2340	3	US-09-022-983-4	Sequence 4, Appli
c 23	71.5	15.6	2477	4	US-09-490-692-3	Sequence 3, Appli
24	70	15.3	2019	2	US-08-245-511-46	Sequence 46, Appl
25	70	15.3	2019	2	US-08-600-993A-46	Sequence 46, Appl
c 26	70	15.3	11864	4	US-08-961-527-61	Sequence 61, Appl
27	67.5	14.7	1386	2	US-08-408-095-22	Sequence 22, Appl
28	67	14.6	4527	3	US-09-045-632-1	Sequence 1, Appli
c 29	67	14.6	6506	4	US-09-453-702B-1	Sequence 1, Appli
c 30	65.5	14.3	6360	4	US-09-171-699-9	Sequence 9, Appli
31	64	13.9	1266	4	US-08-790-186A-3	Sequence 3, Appli
32	64	13.9	1445	4	US-08-949-246-1	Sequence 1, Appli
c 33	63	13.7	1289	4	US-09-149-476-319	Sequence 319, App
c 34	62.5	13.6	576	1	US-08-086-428B-48	Sequence 48, Appl
c 35	62.5	13.6	576	2	US-08-468-570-48	Sequence 48, Appl
c 36	62.5	13.6	576	2	US-08-290-665A-48	Sequence 48, Appl
c 37	62.5	13.6	576	5	PCT-US95-10398-48	Sequence 48, Appl
c 38	62.5	13.6	1302	4	US-09-255-368-5	Sequence 5, Appli
c 39	62.5	13.6	4339	4	US-09-484-970B-164	Sequence 164, App
40	62.5	13.6	5077	2	US-08-687-956A-22	Sequence 22, Appl
c 41	62.5	13.6	49136	4	US-09-422-869-1	Sequence 1, Appli
c 42	62	13.5	628	4	US-08-998-416-87	Sequence 87, Appl
43	62	13.5	1567	3	US-08-961-083-119	Sequence 119, App
44	62	13.5	2557	4	US-08-464-954A-1	Sequence 1, Appli
45	62	13.5	3121	3	US-08-961-083-117	Sequence 117, App

ALIGNMENTS

RESULT 1
US-08-396-479B-7
; Sequence 7, Application US/08396479B
; Patent No. 5612455
; GENERAL INFORMATION:
; APPLICANT: HOEY, Timothy
; TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,479B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59450-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2406 base pairs
; TYPE: nucleic acid

```
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..2337
US-08-396-479B-7

Alignment Scores:
Pred. No.:              5.5e-48        Length:         2406
Score:                  430.00         Matches:        79
Percent Similarity:     96.51%         Conservative:   4
Best Local Similarity:  91.86%         Mismatches:     3
Query Match:            93.68%         Indels:         0
DB:                     1              Gaps:           0

US-09-550-115A-2 (1-86) x US-08-396-479B-7 (1-2406)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 2
US-08-818-823-7
; Sequence 7, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2406 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..2337
US-08-818-823-7

Alignment Scores:
Pred. No.:              5.5e-48        Length:         2406
Score:                  430.00         Matches:        79
Percent Similarity:     96.51%         Conservative:   4
Best Local Similarity:  91.86%         Mismatches:     3
Query Match:            93.68%         Indels:         0
DB:                     1              Gaps:           0

US-09-550-115A-2 (1-86) x US-08-818-823-7 (1-2406)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::|||||||||||:::|||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 3
US-08-396-479B-9
; Sequence 9, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
```

```
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  09:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2647 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..2427
US-08-396-479B-9

Alignment Scores:
Pred. No.:              6.26e-48      Length:          2647
Score:                  430.00        Matches:         79
Percent Similarity:     96.51%        Conservative:    4
Best Local Similarity:  91.86%        Mismatches:      3
Query Match:            93.68%        Indels:          0
DB:                     1             Gaps:            0

US-09-550-115A-2 (1-86) x US-08-396-479B-9 (1-2647)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::||||||||:::|||||||||||||:::|||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428

RESULT 4
US-08-818-823-9
; Sequence 9, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  09:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2647 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..2427
US-08-818-823-9

Alignment Scores:
Pred. No.:              6.26e-48      Length:          2647
Score:                  430.00        Matches:         79
Percent Similarity:     96.51%        Conservative:    4
Best Local Similarity:  91.86%        Mismatches:      3
Query Match:            93.68%        Indels:          0
DB:                     1             Gaps:            0

US-09-550-115A-2 (1-86) x US-08-818-823-9 (1-2647)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::||||||||:::|||||||||||||:::|||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428

RESULT 5
US-08-396-479B-11
; Sequence 11, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
```

```
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 494-8700
;       TELEFAX:  (415) 494-8771
;       TELEX:  210 277299
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3969 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  211..3414
US-08-396-479B-11

Alignment Scores:
Pred. No.:              1.09e-47        Length:         3969
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     1               Gaps:           0

US-09-550-115A-2 (1-86) x US-08-396-479B-11 (1-3969)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::|||||||||:::||||||||||:::|||||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428

RESULT 6
US-08-818-823-11
; Sequence 11, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3969 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..3414
US-08-818-823-11

Alignment Scores:
Pred. No.:              1.09e-47        Length:         3969
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     1               Gaps:           0

US-09-550-115A-2 (1-86) x US-08-818-823-11 (1-3969)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db   1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          |||||||||:::|||||||||:::||||||||||:::|||||||||||||||||||||||
Db   1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
          ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db   1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy     81 TrpSerLysProLysPro 86
          |||||||||||||||||
Db   1411 TGGAGCAAACCAAAGCCT 1428

RESULT 7
PCT-US94-07297-40
; Sequence 40, Application PC/TUS9407297
;  GENERAL INFORMATION:
;    APPLICANT:  Arai, Naoko
;    APPLICANT:  Masuda, Esteban S.
;    APPLICANT:  Tokumitsu, Hiroshi
;    TITLE OF INVENTION:  PURIFIED COMPONENTS OF MAMMALIAN
;    TITLE OF INVENTION:  TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John H. C. Blasdale, Schering-Plough Corporation, M-3-W
;      STREET:  One Giralda Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07940-1000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh IIci
;      OPERATING SYSTEM:  System Software 7.1
```

```
;       SOFTWARE:  Microsoft Word 5.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/07297
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/222,626
;       FILING DATE:  04-APR-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/148,061
;       FILING DATE:  05-NOV-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/113,971
;       FILING DATE:  30-AUG-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/099,998
;       FILING DATE:  30-JUL-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/088,483
;       FILING DATE:  06-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Blasdale, John H. C.
;       REGISTRATION NUMBER:  31,895
;       REFERENCE/DOCKET NUMBER:  DX0392K4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-822-7398
;       TELEFAX:  201-822-7039
;   INFORMATION FOR SEQ ID NO:  40:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4010 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  304..3531
;     FEATURE:
;       NAME/KEY:  unsure
;       LOCATION:  replace(1756..1758, "gta")
;     FEATURE:
;       NAME/KEY:  unsure
;       LOCATION:  replace(1756..1758, "gaa")
;     FEATURE:
;       NAME/KEY:  unsure
;       LOCATION:  replace(1756..1758, "gga")
;     FEATURE:
;       NAME/KEY:  unsure
;       LOCATION:  replace(3090..3092, "agt")
;     FEATURE:
;       NAME/KEY:  unsure
;       LOCATION:  replace(3090..3092, "aga")
;     FEATURE:
;       NAME/KEY:  unsure
;       LOCATION:  replace(3090..3092, "agg")
PCT-US94-07297-40

Alignment Scores:
Pred. No.:              1.1e-47       Length:         4010
Score:                  430.00        Matches:        79
Percent Similarity:     96.51%        Conservative:   4
Best Local Similarity:  91.86%        Mismatches:     3
Query Match:            93.68%        Indels:         0
DB:                     5             Gaps:           0

US-09-550-115A-2 (1-86) x PCT-US94-07297-40 (1-4010)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1264 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1323

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
```

```
Db    1324 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1383

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           ||||||||||:::||||||||||:::||||||||||||||:::||||||||||||||||||||
Db    1384 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1443

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1444 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1503

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1504 TGGAGCAAACCAAAGCCT 1521

RESULT 8
US-08-396-479B-5
; Sequence 5, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2881 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  142..2850
US-08-396-479B-5

Alignment Scores:
Pred. No.:              1.31e-06      Length:         2881
Score:                  122.00        Matches:        34
Percent Similarity:     52.94%        Conservative:   11
Best Local Similarity:  40.00%        Mismatches:     34
Query Match:            26.58%        Indels:         6
DB:                     1             Gaps:           3

US-09-550-115A-2 (1-86) x US-08-396-479B-5 (1-2881)

Qy       2 SerProAlaProPheProPheGlnTyrCysValGluThrAsp---IleProLeuLysThr 20
           ||||||   ||||||   :::                 :::   ||||||   ||||||
Db    1072 TCCCCTGGTCCCTTTGACTATGTGGGGGCCCCACCAGCTGAGAGCATCCCTCAGAAGACA 1131

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
```

```
          |||:::||||||   :::||||||      ||||||         |||      |||
Db   1132 CGGCGGACTTCCAGCGAGCAGGCAGTGGCTCTGCCTCGGTCTGAGGAGCCTGCCTCATGC 1191

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          :::   |||   |||          |||   ||||||           ||||||
Db   1192 AAT---GGGAAGCTGCCCTTGGGAGCAGAGGAGTCTGTGGCTCCTCCAGGAGGTTCC--- 1245

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
                 :::|||:::   :::|||      :::|||:::||||||||||||
Db   1246 ---------CGGAAGGAGGTGGCTGGCATGGACTACCTGGCAGTGCCCTCCCCACTCGCT 1296

Qy     81 TrpSerLysProLys 85
          |||||||||   :::
Db   1297 TGGTCCAAGGCCCGG 1311

RESULT 9
US-08-818-823-5
; Sequence 5, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2881 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  142..2850
US-08-818-823-5

Alignment Scores:
Pred. No.:              1.31e-06       Length:        2881
Score:                  122.00         Matches:       34
Percent Similarity:     52.94%         Conservative:  11
Best Local Similarity:  40.00%         Mismatches:    34
Query Match:            26.58%         Indels:        6
DB:                     1              Gaps:          3

US-09-550-115A-2 (1-86) x US-08-818-823-5 (1-2881)

Qy      2 SerProAlaProPheProPheGlnTyrCysValGluThrAsp---IleProLeuLysThr 20
```

```
          ||||||   ||||||   :::                    :::   ||||||   ||||||
Db   1072 TCCCCTGGTCCCTTTGACTATGTGGGGGCCCCACCAGCTGAGAGCATCCCTCAGAAGACA 1131

Qy     21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
          |||:::||||||   :::||||||      ||||||         |||      |||
Db   1132 CGGCGGACTTCCAGCGAGCAGGCAGTGGCTCTGCCTCGGTCTGAGGAGCCTGCCTCATGC 1191

Qy     41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
          :::   |||   |||          |||   ||||||           ||||||
Db   1192 AAT---GGGAAGCTGCCCTTGGGAGCAGAGGAGTCTGTGGCTCCTCCAGGAGGTTCC--- 1245

Qy     61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
                 :::|||:::   :::|||      :::|||:::||||||||||||
Db   1246 ---------CGGAAGGAGGTGGCTGGCATGGACTACCTGGCAGTGCCCTCCCCACTCGCT 1296

Qy     81 TrpSerLysProLys 85
          |||||||||   :::
Db   1297 TGGTCCAAGGCCCGG 1311

RESULT 10
US-08-396-479B-3
; Sequence 3, Application US/08396479B
; Patent No. 5612455
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,479B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2743 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  240..2390
US-08-396-479B-3

Alignment Scores:
Pred. No.:              1.45e-05       Length:        2743
Score:                  114.00         Matches:       25
Percent Similarity:     61.11%         Conservative:  19
Best Local Similarity:  34.72%         Mismatches:    26
Query Match:            24.84%         Indels:        2
DB:                     1              Gaps:          2

US-09-550-115A-2 (1-86) x US-08-396-479B-3 (1-2743)
```

```
Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::|||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCCGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                  |||:::   :::::::|||            ||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||      |||:::   |||:::|||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442

RESULT 11
US-08-818-823-3
; Sequence 3, Application US/08818823
; Patent No. 5708158
;  GENERAL INFORMATION:
;    APPLICANT:  HOEY, Timothy
;    TITLE OF INVENTION:  NUCLEAR FACTORS AND BINDING ASSAY
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/818,823
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/396,479
;      FILING DATE:  02-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59450-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  210 277299
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2743 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  240..2390
US-08-818-823-3

Alignment Scores:
Pred. No.:              1.45e-05      Length:         2743
Score:                  114.00        Matches:        25
Percent Similarity:     61.11%        Conservative:   19
Best Local Similarity:  34.72%        Mismatches:     26
Query Match:            24.84%        Indels:         2
DB:                     1             Gaps:           2

US-09-550-115A-2 (1-86) x US-08-818-823-3 (1-2743)
```

```
Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::|||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCCGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                  |||:::   :::::::|||            ||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||      |||:::   |||:::|||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442

RESULT 12
US-08-124-981A-1
; Sequence 1, Application US/08124981A
; Patent No. 5837840
;  GENERAL INFORMATION:
;    APPLICANT:  Crabtree, Gerald R.
;    APPLICANT:  No. 5837840throp, Jeffrey P.
;    TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  379 Lytton Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/124,981A
;      FILING DATE:  20-SEP-1993
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Willaim M
;      REGISTRATION NUMBER:  30,223
;      REFERENCE/DOCKET NUMBER:  5490A-226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 326-2400
;      TELEFAX:  (415) 326-2422
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2749 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..2385
US-08-124-981A-1

Alignment Scores:
Pred. No.:              1.45e-05      Length:         2749
Score:                  114.00        Matches:        25
Percent Similarity:     61.11%        Conservative:   19
Best Local Similarity:  34.72%        Mismatches:     26
Query Match:            24.84%        Indels:         2
DB:                     2             Gaps:           2

US-09-550-115A-2 (1-86) x US-08-124-981A-1 (1-2749)

Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::|||||||||:::   :::|||         :::      |||:::
Db    1228 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCCGCCCTCAGTGGCGCTCAAGGTG 1287
```

```
Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1288 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1347

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                   |||:::   :::::::|||           |||||||:::|||:::|||
Db    1348 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1404

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||     |||:::   |||:::|||||||||||||
Db    1405 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1440

RESULT 13
US-09-037-190-45
; Sequence 45, Application US/09037190
; Patent No. 6096515
;  GENERAL INFORMATION:
;    APPLICANT:  Crabtree, Gerald R.
;    APPLICANT:  No. 6096515throp, Jeffrey P.
;    APPLICANT:  Ho, Steffan M.
;    TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,190
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/260,174
;      FILING DATE:  13-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/124,981
;      FILING DATE:  20-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  APV-332.03
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2751 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  240..2387
US-09-037-190-45

Alignment Scores:
Pred. No.:              1.45e-05      Length:         2751
Score:                  114.00        Matches:        25
Percent Similarity:     61.11%        Conservative:   19
Best Local Similarity:  34.72%        Mismatches:     26
Query Match:            24.84%        Indels:         2
DB:                     3             Gaps:           2

US-09-550-115A-2 (1-86) x US-09-037-190-45 (1-2751)
```

```
Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::||||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCAGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         ::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                   |||:::   :::::::|||           |||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||     |||:::   |||:::|||||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442

RESULT 14
US-09-037-192-45
; Sequence 45, Application US/09037192
; Patent No. 6096860
;  GENERAL INFORMATION:
;    APPLICANT:  Crabtree, Gerald R.
;    APPLICANT:  No. 6096860throp, Jeffrey P.
;    APPLICANT:  Ho, Steffan M.
;    TITLE OF INVENTION:  NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/037,192
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/260,174
;      FILING DATE:  13-JUN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/124,981
;      FILING DATE:  20-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  APV-332.04
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2751 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  240..2387
US-09-037-192-45

Alignment Scores:
Pred. No.:              1.45e-05      Length:         2751
Score:                  114.00        Matches:        25
Percent Similarity:     61.11%        Conservative:   19
Best Local Similarity:  34.72%        Mismatches:     26
```

```
Query Match:            24.84%      Indels:        2
DB:                     3           Gaps:          2

US-09-550-115A-2 (1-86) x US-09-037-192-45 (1-2751)

Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::|||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCAGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                  |||:::   :::::::|||            |||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||      |||:::   |||:::||||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442

RESULT 15
US-09-037-143-45
; Sequence 45, Application US/09037143A
; Patent No. 6150099
; GENERAL INFORMATION:
;  APPLICANT: Crabtree, Gerald R.
;  APPLICANT: No. 6150099throp, Jeffrey P.
;  APPLICANT: Ho, Steffan N.
;  TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: APV-332.05
;  CURRENT APPLICATION NUMBER: US/09/037,143A
;  CURRENT FILING DATE: 1998-03-09
;  EARLIER APPLICATION NUMBER:
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 2751
;   TYPE: DNA
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (240)..(2387)
US-09-037-143-45

Alignment Scores:
Pred. No.:              1.45e-05    Length:        2751
Score:                  114.00      Matches:       25
Percent Similarity:     61.11%      Conservative:  19
Best Local Similarity:  34.72%      Mismatches:    26
Query Match:            24.84%      Indels:        2
DB:                     3           Gaps:          2

US-09-550-115A-2 (1-86) x US-09-037-143-45 (1-2751)

Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::|||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCAGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                  |||:::   :::::::|||            |||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||      |||:::   |||:::||||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442
```

Search completed: April 18, 2003, 23:54:38
Job time : 65 secs

GenCore version 5.1.4_p5_4578

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 18, 2003, 22:25:43 ; Search time 118 Seconds
(without alignments)
732.138 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09550115/runat_11042003_141905_18204/app_query.fasta_1.263
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09550115_@CGN_1_1_93_@runat_11042003_141905_18204
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	430	93.7	3969	10	US-09-880-107-2213	Sequence 2213, Ap
2	114	24.8	5502	9	US-10-040-430-45	Sequence 45, Appl
3	114	24.8	5502	9	US-09-349-058-45	Sequence 45, Appl
4	76	16.6	32768	10	US-09-070-927A-17	Sequence 17, Appl
5	71	15.5	32189	9	US-10-079-854-379	Sequence 379, App
6	71	15.5	32189	10	US-09-764-878-379	Sequence 379, App
7	71	15.5	32221	9	US-10-079-854-377	Sequence 377, App
8	71	15.5	32221	10	US-09-764-878-377	Sequence 377, App
c 9	69	15.0	7132	12	US-10-044-090-780	Sequence 780, App
c 10	67.5	14.7	379	9	US-10-046-935-435	Sequence 435, App
c 11	67.5	14.7	379	9	US-09-878-178-435	Sequence 435, App
c 12	67.5	14.7	379	9	US-10-146-502-435	Sequence 435, App
c 13	67	14.6	6506	9	US-10-114-170-1	Sequence 1, Appli
14	66	14.4	463	10	US-09-244-694-87	Sequence 87, Appl
c 15	66	14.4	500	10	US-09-864-761-12394	Sequence 12394, A
c 16	66	14.4	503	9	US-09-736-457-1364	Sequence 1364, Ap
c 17	66	14.4	503	9	US-09-902-941-1364	Sequence 1364, Ap
c 18	66	14.4	503	9	US-09-849-626-1364	Sequence 1364, Ap
c 19	66	14.4	503	9	US-10-017-754-1364	Sequence 1364, Ap
20	66	14.4	1214	10	US-09-764-864-87	Sequence 87, Appl
21	65.5	14.3	2567	9	US-09-925-299-167	Sequence 167, App
22	65.5	14.3	2567	10	US-09-925-299-167	Sequence 167, App
23	64.5	14.1	530	9	US-09-796-692-9499	Sequence 9499, Ap
24	64.5	14.1	1887	10	US-09-954-456-1624	Sequence 1624, Ap
25	64.5	14.1	2085	9	US-09-405-920-1	Sequence 1, Appli
26	64.5	14.1	6663	10	US-09-037-657-28	Sequence 28, Appl
27	64.5	14.1	11832	10	US-09-037-657-38	Sequence 38, Appl
28	64	13.9	1654	9	US-10-108-605-30	Sequence 30, Appl
29	64	13.9	2076	10	US-09-823-847-1	Sequence 1, Appli
30	64	13.9	6339	10	US-09-727-770-3	Sequence 3, Appli
31	63	13.7	550	10	US-09-560-863-463	Sequence 463, App
32	63	13.7	766	10	US-09-910-943-489	Sequence 489, App
33	63	13.7	1023	10	US-09-822-849A-537	Sequence 537, App
c 34	63	13.7	1289	9	US-09-809-391-319	Sequence 319, App
c 35	63	13.7	1817	10	US-09-764-864-457	Sequence 457, App
c 36	63	13.7	32195	9	US-10-092-154-1512	Sequence 1512, Ap
c 37	63	13.7	32195	10	US-09-764-847-1512	Sequence 1512, Ap
38	62.5	13.6	400	7	US-08-781-986A-1158	Sequence 1158, Ap
c 39	62.5	13.6	942	9	US-09-938-842A-1789	Sequence 1789, Ap
40	62.5	13.6	1289	9	US-09-989-442-55	Sequence 55, Appl
41	62.5	13.6	1289	9	US-10-073-865-50	Sequence 50, Appl
42	62.5	13.6	1289	10	US-09-764-853-240	Sequence 240, App
c 43	62.5	13.6	1302	9	US-09-866-248A-5	Sequence 5, Appli
c 44	62.5	13.6	1739	10	US-09-731-872-225	Sequence 225, App
c 45	62.5	13.6	49136	10	US-09-768-877-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-880-107-2213
; Sequence 2213, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2213
; LENGTH: 3969
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L41067
US-09-880-107-2213

Alignment Scores:

```
Pred. No.:              1.03e-50        Length:         3969
Score:                  430.00          Matches:        79
Percent Similarity:     96.51%          Conservative:   4
Best Local Similarity:  91.86%          Mismatches:     3
Query Match:            93.68%          Indels:         0
DB:                     10              Gaps:           0

US-09-550-115A-2 (1-86) x US-09-880-107-2213 (1-3969)

Qy       1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
           |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    1171 CTTGGCCCTGCAGTTTTTCCATTTCAGTACTGTGTAGAGACTGACATCCCTCTCAAAACA 1230

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    1231 AGGAAAACTTCTGAAGATCAAGCTGCCATACTACCAGGAAAATTAGAGCTGTGTTCAGAT 1290

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           |||||||||:::|||||||||:::||||||||||||:::||||||||||||||||||||
Db    1291 GACCAAGGGAGTTTATCACCAGCCCGGGAGACTTCAATAGATGATGGCCTTGGATCTCAG 1350

Qy      61 TyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSerProPheThr 80
           ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1351 TATCCTTTAAAGAAAGATTCATGTGGTGATCAGTTTCTTTCAGTTCCTTCACCCTTTACC 1410

Qy      81 TrpSerLysProLysPro 86
           |||||||||||||||||
Db    1411 TGGAGCAAACCAAAGCCT 1428

RESULT 2
US-10-040-430-45
; Sequence 45, Application US/10040430
; Publication No. US20030049641A1
; GENERAL INFORMATION:
;  APPLICANT: Crabtree, Gerald R.
;  APPLICANT:  No..US20030049641Althrop, Jeffrey P.
;  APPLICANT:  Ho, Steffan M.
;  APPLICANT:  Flanagan, William M.
;  TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES AND SCREENING
;  TITLE OF INVENTION:  METHODS FOR IMMUNOSUPPRESSIVE AGENTS
;  FILE REFERENCE: APV-008.04
;  CURRENT APPLICATION NUMBER: US/10/040,430
;  CURRENT FILING DATE:  2002-01-07
;  PRIOR APPLICATION NUMBER: US/09/232,346
;  PRIOR FILING DATE: 1999-01-15
;  PRIOR APPLICATION NUMBER: 08/507,032
;  PRIOR FILING DATE: 1995-07-31
;  PRIOR APPLICATION NUMBER: 08/228,944
;  PRIOR FILING DATE: 1994-04-18
;  PRIOR APPLICATION NUMBER: 07/749,385
;  PRIOR FILING DATE: 1991-08-22
;  PRIOR APPLICATION NUMBER: 08/260,174
;  PRIOR FILING DATE: 1994-06-13
;  PRIOR APPLICATION NUMBER: 08/124,981
;  PRIOR FILING DATE: 1993-09-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 5502
;   TYPE: DNA
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (240)..(2387)
US-10-040-430-45

Alignment Scores:
Pred. No.:              6.64e-06        Length:         5502
Score:                  114.00          Matches:        25
Percent Similarity:     61.11%          Conservative:   19
Best Local Similarity:  34.72%          Mismatches:     26
Query Match:            24.84%          Indels:         2
DB:                     9               Gaps:           2

US-09-550-115A-2 (1-86) x US-10-040-430-45 (1-5502)

Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCAGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                   |||:::   ::::::|||           ||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||      |||:::   |||:::||||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442

RESULT 3
US-09-349-058-45
; Sequence 45, Application US/09349058
; Publication No. US20030056229A1
; GENERAL INFORMATION:
;  APPLICANT: Crabtree, Gerald R.
;  APPLICANT: No. US20030056229Althrop, Jeffrey P.
;  APPLICANT: Ho, Steffan N.
;  TITLE OF INVENTION: NF-AT POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: APV-332.12
;  CURRENT APPLICATION NUMBER: US/09/349,058
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: 08/260,174
;  EARLIER FILING DATE: 1994-06-13
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 5502
;   TYPE: DNA
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (240)..(2387)
US-09-349-058-45

Alignment Scores:
Pred. No.:              6.64e-06        Length:         5502
Score:                  114.00          Matches:        25
Percent Similarity:     61.11%          Conservative:   19
Best Local Similarity:  34.72%          Mismatches:     26
Query Match:            24.84%          Indels:         2
DB:                     9               Gaps:           2

US-09-550-115A-2 (1-86) x US-09-349-058-45 (1-5502)

Qy      16 IleProLeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu 35
           :::|||:::|||:::||||||||:::   :::|||         :::      |||:::
Db    1230 GTCCCTGTCAAGTCCCGCAAGACCACCCTGGAGCAGCAGCCCTCAGTGGCGCTCAAGGTG 1289

Qy      36 GluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAsp 55
           |||         :::|||   |||:::      |||      :::   :::   :::|||
Db    1290 GAGCCCGTCGGGGAGGACCTGGGCAGCCCCCCGCCCCCGGCCGACTTCGCGCCCGAAGAC 1349

Qy      56 GlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerVal 75
                   |||:::   ::::::|||           ||||||:::|||:::|||
Db    1350 TACTCCTCTTTCCAGCAC---ATCAGGAAGGGCGGCTTCTGCGACCAGTACCTGGCGGTG 1406

Qy      76 ProSer---ProPheThrTrpSerLysProLysPro 86
           |||      |||:::   |||:::||||||||||||
Db    1407 CCGCAGCACCCCTACCAGTGGGCGAAGCCCAAGCCC 1442

RESULT 4
US-09-070-927A-17
; Sequence 17, Application US/09070927A
```

```
; Patent No. US20020120116A1
;  GENERAL INFORMATION:
;       APPLICANT: Charles A. Kunsch
;                  Patrick J. Dillon
;                  Steven Barash
;       TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
;       NUMBER OF SEQUENCES: 982
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/070,927A
;            FILING DATE: 04-May-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/046,655
;            FILING DATE: 1997-05-16
;            APPLICATION NUMBER: 60/044,031
;            FILING DATE: 1997-05-06
;            APPLICATION NUMBER: 60/066,009
;            FILING DATE: 1997-11-14
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kenley K. Hoover
;            REGISTRATION NUMBER: 40,302
;            REFERENCE/DOCKET NUMBER: PB369
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (301) 309-8504
;            TELEFAX: (301) 309-8512
;    INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 32768 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-070-927A-17

Alignment Scores:
Pred. No.:              15.4          Length:        32768
Score:                  76.00         Matches:       23
Percent Similarity:     48.28%        Conservative:  19
Best Local Similarity:  26.44%        Mismatches:    27
Query Match:            16.56%        Indels:        18
DB:                     10            Gaps:          4

US-09-550-115A-2 (1-86) x US-09-070-927A-17 (1-32768)

Qy       3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
           |||      :::      :::      |||||||||:::      ||||||   :::|||
Db      65 CCGAACACTTACACGGTCAAAWTTGATGTAGAAACCGAGGGTTCTTTAAAAATCAAAAAA 124

Qy      23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAspGln 42
             ::::::      ||||||:::|||||||         :::          |||
Db     125 GTTGATAAAGAATCAGGTGCTATTGTACCAGGA------ACGGTTTTCCATTTAGACTTT 178

Qy      43 GlyAsnLeuSerProSerArgGluThrSerValAsp--------------------Asp 55
           |||         |||:::::::::   :::   |||                     |||
Db     179 GGAAAAACTTTACCTGCAAAAGACGTGACTACTGACAAAGAAGGCATTGCTACATTGGAT 238

Qy      56 GlyLeu------GlySerGlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeu 73
           |||:::      |||:::::::      :::   :::
Db     239 GGGATTCCTCATGGTACAAAGGTAACCATTACCGAAAAA---------------------- 277

Qy      74 SerValProSerProPheThr 80
```

```
           |||||||||:::|||:::|||
Db     278 TCGGTGCCGGCGCCTTATACG 298

RESULT 5
US-10-079-854-379
; Sequence 379, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121C1
;  CURRENT APPLICATION NUMBER: US/10/079,854
;  CURRENT FILING DATE:  2002-02-22
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 379
;   LENGTH: 32189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-079-854-379

Alignment Scores:
Pred. No.:              76.6          Length:        32189
Score:                  71.00         Matches:       19
Percent Similarity:     52.83%        Conservative:  9
Best Local Similarity:  35.85%        Mismatches:    15
Query Match:            15.47%        Indels:        10
DB:                     9             Gaps:          3

US-09-550-115A-2 (1-86) x US-10-079-854-379 (1-32189)

Qy      32 ProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThr 51
           ||||||:::      :::||||||  :::|||         :::|||  ||||||
Db   25457 CCAGGAAGGGCATTTGTGTGTTCTGGGGAGCAGCTGCGGAGAGCTCCCCAGAGGGAAGAA 25516

Qy      52 SerValAspAspGlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGln 71
           :::      :::|||||||  :::|||  |||                   |||
Db   25517 GCA------GAAGGACTTGAGGCCCAAGCACCT-----------------GGTGCTAGT 25552

Qy      72 PheLeuSerValProSerProPhe------ThrTrpSer 82
             |||   :::|||:::   |||      |||||||||
Db   25553 CCTCTGGACATCCCAGCCACTTTTGCTGCGACTTGGTCC 25591

RESULT 6
US-09-764-878-379
; Sequence 379, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121
;  CURRENT APPLICATION NUMBER: US/09/764,878
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 379
;   LENGTH: 32189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-878-379

Alignment Scores:
Pred. No.:              76.6          Length:        32189
Score:                  71.00         Matches:       19
Percent Similarity:     52.83%        Conservative:  9
Best Local Similarity:  35.85%        Mismatches:    15
Query Match:            15.47%        Indels:        10
DB:                     10            Gaps:          3

US-09-550-115A-2 (1-86) x US-09-764-878-379 (1-32189)
```

```
Qy      32 ProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThr 51
           |||||||:::      :::|||||||   :::|||         :::|||   ||||||
Db  25457 CCAGGAAGGGCATTTGTGTGTTCTGGGGAGCAGCTGCGGAGAGCTCCCCAGAGGGAAGAA 25516

Qy      52 SerValAspAspGlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGln 71
           :::      :::|||||   :::|||   |||                   |||
Db  25517 GCA------GAAGGACTTGAGGCCCAAGCACCT-----------------GGTGCTAGT 25552

Qy      72 PheLeuSerValProSerProPhe------ThrTrpSer 82
              |||   :::|||:::   |||      |||||||||
Db  25553 CCTCTGGACATCCCAGCCACTTTTGCTGCGACTTGGTCC 25591

RESULT 7
US-10-079-854-377
; Sequence 377, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121C1
;  CURRENT APPLICATION NUMBER: US/10/079,854
;  CURRENT FILING DATE:  2002-02-22
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 377
;   LENGTH: 32221
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (7464)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-079-854-377

Alignment Scores:
Pred. No.:              76.7          Length:        32221
Score:                  71.00         Matches:       19
Percent Similarity:     52.83%        Conservative:  9
Best Local Similarity:  35.85%        Mismatches:    15
Query Match:            15.47%        Indels:        10
DB:                     9             Gaps:          3

US-09-550-115A-2 (1-86) x US-10-079-854-377 (1-32221)

Qy      32 ProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThr 51
           ||||||:::      :::|||||||   :::|||         :::|||   ||||||
Db  25488 CCAGGAAGGGCATTTGTGTGTTCTGGGGAGCAGCTGCGGAGAGCTCCCCAGAGGGAAGAA 25547

Qy      52 SerValAspAspGlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGln 71
           :::      :::|||||   :::|||   |||                   |||
Db  25548 GCA------GAAGGACTTGAGGCCCAAGCACCT-----------------GGTGCTAGT 25583

Qy      72 PheLeuSerValProSerProPhe------ThrTrpSer 82
              |||   :::|||:::   |||      |||||||||
Db  25584 CCTCTGGACATCCCAGCCACTTTTGCTGCGACTTGGTCC 25622

RESULT 8
US-09-764-878-377
; Sequence 377, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121
;  CURRENT APPLICATION NUMBER: US/09/764,878
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 377
;   LENGTH: 32221
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (7464)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-878-377

Alignment Scores:
Pred. No.:              76.7          Length:        32221
Score:                  71.00         Matches:       19
Percent Similarity:     52.83%        Conservative:  9
Best Local Similarity:  35.85%        Mismatches:    15
Query Match:            15.47%        Indels:        10
DB:                     10            Gaps:          3

US-09-550-115A-2 (1-86) x US-09-764-878-377 (1-32221)

Qy      32 ProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeuSerProSerArgGluThr 51
           ||||||:::      :::|||||||   :::|||         :::|||   ||||||
Db  25488 CCAGGAAGGGCATTTGTGTGTTCTGGGGAGCAGCTGCGGAGAGCTCCCCAGAGGGAAGAA 25547

Qy      52 SerValAspAspGlyLeuGlySerGlnTyrProLeuLysLysAspSerSerGlyAspGln 71
           :::      :::|||||   :::|||   |||                   |||
Db  25548 GCA------GAAGGACTTGAGGCCCAAGCACCT-----------------GGTGCTAGT 25583

Qy      72 PheLeuSerValProSerProPhe------ThrTrpSer 82
              |||   :::|||:::   |||      |||||||||
Db  25584 CCTCTGGACATCCCAGCCACTTTTGCTGCGACTTGGTCC 25622

RESULT 9
US-10-044-090-780/c
; Sequence 780, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 780
;   LENGTH: 7132
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 198499.13
US-10-044-090-780

Alignment Scores:
Pred. No.:              21            Length:        7132
Score:                  69.00         Matches:       13
Percent Similarity:     51.22%        Conservative:  8
Best Local Similarity:  31.71%        Mismatches:    6
Query Match:            15.03%        Indels:        14
DB:                     12            Gaps:          1

US-09-550-115A-2 (1-86) x US-10-044-090-780 (1-7132)

Qy       7 ProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLysThrSerGluAsp 26
           |||   |||   |||||||:::   |||:::|||||||:::|||
Db   3561 CCAACCCAGGTCTGTGTTGATGAGGACCTTCCGCTCAGAACC----------------- 3520

Qy      27 GlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAspAspGlnGlyAsnLeuSer 46
                                    ::::::::::|||      |||   ||||||::::::
Db   3519 ------------------------ATCCAAGTTTGCCCGTTCGATCTTGGCAATGTGGCC 3484

Qy      47 Pro 47
           |||
Db   3483 CCA 3481

RESULT 10
```

```
US-10-046-935-435/c
; Sequence 435, Application US/10046935
; Patent No. US20020156011A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.527C1
;  CURRENT APPLICATION NUMBER: US/10/046,935
;  CURRENT FILING DATE:  2002-01-15
;  NUMBER OF SEQ ID NOS: 2239
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 435
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-046-935-435

Alignment Scores:
Pred. No.:              0.775          Length:         379
Score:                  67.50          Matches:        28
Percent Similarity:     42.53%         Conservative:   9
Best Local Similarity:  32.18%         Mismatches:     41
Query Match:            14.71%         Indels:         9
DB:                     9              Gaps:           4

US-09-550-115A-2 (1-86) x US-10-046-935-435 (1-379)

Qy       3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
           |||      |||||||||   |||                              |||
Db     361 CCTCAAACTTTTCCTTTTTCTTAT--------------AGAAAATGGAAAGCTTACAAT 317

Qy      23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu---GluIleCys------Ser 39
           ||||||      :::|||||||   |||         |||      :::|||
Db     316 ACCTCCTCCATCAAAGCGGCAGGCCTACGAGCCAGCCTGAACAGGGTTTGCCTTGGAAAA 257

Qy      40 AspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySer 59
           |||             |||   |||                |||:::   |||:::|||
Db     256 GATGTGGCCTGAGGTTTAGAGCCGCTTTGTGCGGGGATGGTGGAGGCTGGGGTGGGGGTG 197

Qy      60 GlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSer---Pro 78
           :::         ::::::         |||      ||||||   :::   |||   |||
Db     196 AGAAAAGGGAGAAGGCGGAAGGGGGACGGACAGTTCTTTCTTTTTCTCTCTAGCTTACCC 137

Qy      79 PheThrTrpSerLysProLys 85
           |||      :::||||||||||
Db     136 TTTTTTCTAAATAAGCCCAAA 116

RESULT 11
US-09-878-178-435/c
; Sequence 435, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.527
;  CURRENT APPLICATION NUMBER: US/09/878,178
;  CURRENT FILING DATE:  2001-06-08
;  NUMBER OF SEQ ID NOS: 2237
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 435
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-878-178-435

Alignment Scores:
Pred. No.:              0.775          Length:         379
Score:                  67.50          Matches:        28
Percent Similarity:     42.53%         Conservative:   9
Best Local Similarity:  32.18%         Mismatches:     41
Query Match:            14.71%         Indels:         9
DB:                     9              Gaps:           4

US-09-550-115A-2 (1-86) x US-09-878-178-435 (1-379)

Qy       3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
           |||      |||||||||   |||                              |||
Db     361 CCTCAAACTTTTCCTTTTTCTTAT--------------AGAAAATGGAAAGCTTACAAT 317

Qy      23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu---GluIleCys------Ser 39
           ||||||      :::|||||||   |||         |||      :::|||
Db     316 ACCTCCTCCATCAAAGCGGCAGGCCTACGAGCCAGCCTGAACAGGGTTTGCCTTGGAAAA 257

Qy      40 AspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySer 59
           |||             |||   |||                |||:::   |||:::|||
Db     256 GATGTGGCCTGAGGTTTAGAGCCGCTTTGTGCGGGGATGGTGGAGGCTGGGGTGGGGGTG 197

Qy      60 GlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSer---Pro 78
           :::         ::::::         |||      ||||||   :::   |||   |||
Db     196 AGAAAAGGGAGAAGGCGGAAGGGGGACGGACAGTTCTTTCTTTTTCTCTCTAGCTTACCC 137

Qy      79 PheThrTrpSerLysProLys 85
           |||      :::||||||||||
Db     136 TTTTTTCTAAATAAGCCCAAA 116

RESULT 12
US-10-146-502-435/c
; Sequence 435, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.527C2
;  CURRENT APPLICATION NUMBER: US/10/146,502
;  CURRENT FILING DATE:  2002-05-14
;  NUMBER OF SEQ ID NOS: 2241
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 435
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-146-502-435

Alignment Scores:
Pred. No.:              0.775          Length:         379
Score:                  67.50          Matches:        28
Percent Similarity:     42.53%         Conservative:   9
Best Local Similarity:  32.18%         Mismatches:     41
Query Match:            14.71%         Indels:         9
DB:                     9              Gaps:           4

US-09-550-115A-2 (1-86) x US-10-146-502-435 (1-379)

Qy       3 ProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThrArgLys 22
           |||      |||||||||   |||                              |||
Db     361 CCTCAAACTTTTCCTTTTTCTTAT--------------AGAAAATGGAAAGCTTACAAT 317

Qy      23 ThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeu---GluIleCys------Ser 39
           ||||||      :::|||||||   |||         |||      :::|||
Db     316 ACCTCCTCCATCAAAGCGGCAGGCCTACGAGCCAGCCTGAACAGGGTTTGCCTTGGAAAA 257

Qy      40 AspAspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySer 59
           |||             |||   |||                |||:::   |||:::|||
```

```
Db    256 GATGTGGCCTGAGGTTTAGAGCCGCTTTGTGCGGGGATGGTGGAGGCTGGGGTGGGGGTG 197

Qy     60 GlnTyrProLeuLysLysAspSerSerGlyAspGlnPheLeuSerValProSer---Pro 78
          :::         ::::::      |||      ||||||   :::   |||   |||
Db    196 AGAAAAGGGAGAAGGCGGAAGGGGGACGGACAGTTCTTTCTTTTTCTCTCTAGCTTACCC 137

Qy     79 PheThrTrpSerLysProLys 85
          |||      :::|||||||||
Db    136 TTTTTTCTAAATAAGCCCAAA 116

RESULT 13
US-10-114-170-1/c
; Sequence 1, Application US/10114170
; Publication No. US20030023075A1
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. US20030023075Alel Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/114,170
;             FILING DATE: 01-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/453,702
;             FILING DATE: 03-DEC-1999
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 6506
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-114-170-1

Alignment Scores:
Pred. No.:              35.7          Length:        6506
Score:                  67.00         Matches:       27
Percent Similarity:     36.36%        Conservative:  9
Best Local Similarity:  27.27%        Mismatches:    23
Query Match:            14.60%        Indels:        40
DB:                     9             Gaps:          4

US-09-550-115A-2 (1-86) x US-10-114-170-1 (1-6506)

Qy     18 LeuLysThrArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIle 37
          |||||||||      ||||||   |||                          |||||||||
Db   4352 TTAAAAACCGGTTCAACCTCTCATGAC-----------------------TTAGAAATT 4317

Qy     38 CysSerAspAspGlnGlyAsnLeuSerProSerArgGluThr----------------- 51
                |||      ||||||   :::||||||   |||
Db   4316 CAGTGTGAC------GGAAATACAGCCCCAAGTATCGAAGTACTAGCATCAGAGTTTGAT 4263

Qy     52 ----------------------------------------------------SerVal 53
                                                              :::
Db   4262 GCTAATAACTTTACGCTGAATAAAGGAACTGCTGAGAACATTGGCTTTGCCATTTTCCTC 4203

Qy     54 AspAspGlyLeu-----------------GlySerGlnTyrProLeuLysLysAspSer 67
          |||||||||                  |||   :::   |||||||||   :::
Db   4202 GATGATGGGACTTTGACCAGAATGAGACCAGGGTTAAAAACACCGCTGAAAGATAATGGC 4143

Qy     68 SerGlyAspGlnPheLeuSerValProSerProPheThrTrpSerLysProLysPro 86
          ||||||      ||||||::::::   :::   :::      :::         |||
Db   4142 TCTGGTCAATATTTTCTCAATTTAAATGCACAATATGCCCGCACAAGTGGCAATCCT 4086

RESULT 14
US-09-244-694-87
; Sequence 87, Application US/09244694
; Patent No. US20020026037A1
; GENERAL INFORMATION:
;  APPLICANT: Hu, Jing-Shan
;  APPLICANT: Olsen, Henrik S.
;  APPLICANT: Rosen, Craig A.
;  TITLE OF INVENTION: Vascular Endothelial Growth Factor 3
;  FILE REFERENCE: 1488.1040003
;  CURRENT APPLICATION NUMBER: US/09/244,694
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: US 09/132,088
;  EARLIER FILING DATE: 1998-08-10
;  EARLIER APPLICATION NUMBER: US 09/033,662
;  EARLIER FILING DATE: 1998-03-03
;  EARLIER APPLICATION NUMBER: US 08/469,641
;  EARLIER FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 194
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 87
;   LENGTH: 463
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-244-694-87

Alignment Scores:
Pred. No.:              1.63          Length:        463
Score:                  66.00         Matches:       28
Percent Similarity:     43.12%        Conservative:  19
Best Local Similarity:  25.69%        Mismatches:    34
Query Match:            14.38%        Indels:        29
DB:                     10            Gaps:          5

US-09-550-115A-2 (1-86) x US-09-244-694-87 (1-463)

Qy      1 LeuSerProAlaProPheProPheGlnTyrCysValGluThrAspIleProLeuLysThr 20
          :::||||||         |||      :::|||      :::      |||   :::|||
Db     67 ATCTCCCCAGTCAAACTCCCCACCTGGCACTGTCCCCGGAGCGGGTCGCCCCGCCGCACG 126

Qy     21 Arg-----------------LysThrSerGluAspGlnAlaAlaIleLeuProGlyLys 34
          |||                 ::::::          :::              |||   :::
Db    127 CGCGGACGGACGCTTGGCGCCAGAAGCGAGAGCCCTCGGGTGTTCCCCCCGCCTCACCGG 186

Qy     35 LeuGluIleCysSerAspAspGlnGlyAsnLeuSerPro---SerArgGluThrServal 53
          :::         :::|||:::         :::|||||||   :::||||||||||   |||
Db    187 GTCAGTGA-CTCAACGATGAGAGTAGTGGTATTTCACCGGCGGCCCGCGAGACCGGCGTG 245

Qy     54 --------------AspAspGlyLeuGlySerGlnTyrPro----------------- 62
                       :::||||||   |||   :::   |||
Db    246 CCCCGACCCCGACGCGAGGACGGGGCCGGGCGCCGGGGGCCTCACACTTATTCTACACCT 305

Qy     63 ----------------------LeuLysLysAspSerSerGlyAspGlnPheLeuSer 74
                               |||:::      |||:::|||      |||   :::
Db    306 CTCATGTCTCTTCACCGTGCCAGACTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCT 365
```

```
Qy      75 -Val---ProSerProPheThrTrp 81
           :::   ||||||||||||   |||
Db     366 GATTCCGCCAAGCCCGTTCCCTTGG 390

RESULT 15
US-09-864-761-12394/c
; Sequence 12394, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12394
;   LENGTH: 500
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL121585.13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.71
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.63
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.46
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.59
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.6
US-09-864-761-12394

Alignment Scores:
Pred. No.:              1.8          Length:         500
Score:                  66.00        Matches:        22
Percent Similarity:     42.47%       Conservative:   9
Best Local Similarity:  30.14%       Mismatches:     24
Query Match:            14.38%       Indels:         18
DB:                     10           Gaps:           2

US-09-550-115A-2 (1-86) x US-09-864-761-12394 (1-500)

Qy      21 ArgLysThrSerGluAspGlnAlaAlaIleLeuProGlyLysLeuGluIleCysSerAsp 40
           :::||||||||||:::      |||||||:::|||   |||||||||
Db     255 AAAAAGACATCTAAAAAACTGGCAGCCCTTCTGCAGGGCAAGCTGACATCCGCACACCAG 196

Qy      41 AspGlnGlyAsnLeuSerProSerArgGluThrSerValAspAspGlyLeuGlySerGln 60
           :::      :::||||||         |||          :::|||   |||   |||
Db     195 GAGAATACCAGCCTGAGTGAGGAGGAGGAGAGGAAGTGAGGAGAAGGAAGGGGAGGACAG 136

Qy      61 ---TyrProLeuLysLysAspSerSer------------------------------- 68
              :::   |||   :::   |||:::
Db     135 ACGTTCACACTGCCACGTATGTCTACGTGGATTTTTGGTTTTCAGCTTCCCCCACCCCAC 76

Qy      69 ------------------GlyAspGlnPheLeuSerVal 75
                             |||||||||   ||||||
Db      75 TGGCTCTTCTTAATTAGAAGTGACCAGTTCACCTCTGTG 37


Search completed: April 18, 2003, 23:57:13
Job time : 150 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:20:32 ; Search time 36 Seconds
(without alignments)
318.321 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	430	93.7	708	17	AAW02251	Human transcriptio
2	430	93.7	1075	16	AAR75353	Human NF-AT120, X
3	122	26.6	902	17	AAW02250	Human transcriptio
4	122	26.6	960	22	AAU32449	Novel human secret
5	115	25.1	902	20	AAW85730	NF-AT3. Mus muscu
6	114	24.8	716	16	AAR66874	Human NF-AT120, C
7	114	24.8	716	16	AAR70357	NF-AT transcriptio
8	114	24.8	716	17	AAW02249	Human transcriptio
9	114	24.8	716	21	AAY96535	Human nuclear fact
10	114	24.8	716	21	AAY96565	Human nuclear fact
11	114	24.8	716	22	AAB90774	Human shear stress
12	114	24.8	716	22	AAB66497	Human NF-AT protei
13	114	24.8	716	22	AAB66501	Human NF-ATc prote
14	114	24.8	716	23	AAE21559	Human cytoplasmic
15	114	24.8	716	23	AAE21563	Human cytoplasmic
16	114	24.8	716	23	AAE21566	Human cytoplasmic
17	114	24.8	761	16	AAR70358	Human NF-AT transc
18	114	24.8	897	22	ABG14209	Novel human diagno
19	77	16.8	1236	22	ABB65297	Drosophila melanog
20	77	16.8	1257	22	ABB67273	Drosophila melanog
21	71.5	15.6	418	22	ABG16205	Novel human diagno
22	71.5	15.6	418	22	ABG16613	Novel human diagno
23	71.5	15.6	418	22	ABG17215	Novel human diagno
24	71.5	15.6	513	21	AAB19771	Human dentin matri
25	71.5	15.6	936	22	ABG16765	Novel human diagno
26	70	15.3	643	16	AAR70152	Streptococcus pneu
27	68.5	14.9	481	22	ABG28634	Novel human diagno
28	68	14.8	581	20	AAY42774	Rat neuronal immed
29	68	14.8	581	22	AAM51819	Rat apoptase L100
30	67.5	14.7	293	16	AAR81703	Human NF-AT transc
31	67.5	14.7	298	16	AAR81705	Human NF-AT transc
32	67.5	14.7	314	17	AAW00998	Apoptosis-regulati
33	67.5	14.7	314	20	AAW94324	Human Nip2 protein
34	67	14.6	80	22	AAM84385	Human immune/haema
35	67	14.6	1112	19	AAW73061	Rat GRIP. Rattus
36	66.5	14.5	289	22	AAB65676	Novel protein kina
37	66	14.4	172	22	AAM82398	Human immune/haema
38	66	14.4	430	20	AAW96272	Testis specific is
39	66	14.4	430	22	ABG19615	Novel human diagno
40	66	14.4	728	13	AAR20100	Contains alpha- an
41	66	14.4	856	21	AAB41660	Human ORFX ORF1424
42	66	14.4	1173	22	AAB97911	Human G-protein ac
43	65.5	14.3	166	21	AAY50924	Human fetal brain
44	65.5	14.3	456	23	ABB89424	Human polypeptide
45	65.5	14.3	557	21	AAB53400	Human colon cancer

ALIGNMENTS

RESULT 1
AAW02251
ID AAW02251 standard; Protein; 708 AA.
XX
AC AAW02251;
XX
DT 18-NOV-1996 (first entry)
XX
DE Human transcription factor NFAT4a.
XX
KW Nuclear factor of activated T-cells; NFAT; NFAT4a;
KW transcription factor; cytokine; gene expression; binding assay;
KW immune system disease; therapy; diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 1..699
FT /note= "N-terminal sequence shared by
FT NFAT4a, NFAT4b and NFAT4c"
FT Domain 411..702
FT /label= Rel_domain
XX
PN WO9626959-A1.
XX
PD 06-SEP-1996.
XX
PF 04-MAR-1996; 96WO-US03113.
XX
PR 02-MAR-1995; 95US-0396479.
XX
PA (TULA-) TULARIK INC.
XX

```
PI   Hoey T;
XX
DR   WPI; 1996-412738/41.
DR   N-PSDB; AAT33677.
XX
PT   DNA mol. encoding human nuclear factors of activated T cells -
PT   useful for screening potential therapeutic and diagnostic agents for
PT   immune system diseases
XX
PS   Claim 6; Page 52-54; 64pp; English.
XX
CC   4 Types of human nuclear factor of activated T-cells class 4,
CC   NFAT4a (AAW02251), NFAT4b and NFAT4c, result from alternative
CC   splicing downstream of the rel homology domain.  The 3 types
CC   have identical N-terminal sequences, but C-terminal sequences
CC   differ for NFATb (AAW02252) and NFATc (AAW02253) from that for
CC   NFATa.  NFATs (see also AAW02248-50) include regulators of cytokine
CC   gene expression that modulate immune system function.  Recombinant
CC   NFATs, or NFAT fragments contg. the rel domain, can be expressed in
CC   prokaryotic or eukaryotic host cells.  They are used in high-
CC   throughput screenings to identify agents useful in the diagnosis or
CC   treatment of diseases associated with expression of a gene modulated
CC   by a transcription complex contg. NFAT(s).
XX
SQ   Sequence   708 AA;

  Query Match             93.7%;  Score 430;  DB 17;  Length 708;
  Best Local Similarity   91.9%;  Pred. No. 3.2e-46;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || ||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
Db     381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 2
AAR75353
ID   AAR75353 standard; Protein; 1075 AA.
XX
AC   AAR75353;
XX
DT   12-AUG-1995  (first entry)
XX
DE   Human NF-AT120, X subfamily.
XX
KW   NF-AT120; nuclear factor of activated T cells 120 protein;
KW   transcription; differentiation; cytokine; T-cell; T-lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO9502053-A.
XX
PD   19-JAN-1995.
XX
PF   05-JUL-1994;   94WO-US07297.
XX
PR   06-JUL-1993;   93US-0088483.
PR   30-JUL-1993;   93US-0099998.
PR   30-AUG-1993;   93US-0113971.
PR   05-NOV-1993;   93US-0148061.
PR   04-APR-1994;   94US-0222626.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Arai N,  Masuda ES,  Tokumitsu H;
XX
DR   WPI; 1995-066898/09.
DR   N-PSDB; AAQ84303.
XX
PT   Nuclear Factor of Activated T cells 120 protein, antibody and
PT   nucleic acid - useful as transcriptional regulator of genes, such
PT   as cytokines
XX
PS   Disclosure; Page 78-83; 100pp; English.
XX
CC   A DNA fragment amplified from Jurkat cDNA by PCR had the sequence
CC   given in AAQ84300 and encoded NF-AT120 (Nuclear Factor of
CC   Activated T cells protein 120) (AAR66873).  Amplified DNA was used
CC   to screen for related NF-AT proteins by hybridization, leading to
CC   the discovery of 3 related subfamilies, designated class C, P and
CC   X (AAR66874, AAR66875, AAR75353).
XX
SQ   Sequence   1075 AA;

  Query Match             93.7%;  Score 430;  DB 16;  Length 1075;
  Best Local Similarity   91.9%;  Pred. No. 5.6e-46;
  Matches   79;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || ||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
Db     321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||| ||||||||||||||||||
Db     381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 3
AAW02250
ID   AAW02250 standard; Protein; 902 AA.
XX
AC   AAW02250;
XX
DT   17-NOV-1996  (first entry)
XX
DE   Human transcription factor NFAT3.
XX
KW   Nuclear factor of activated T-cells; NFAT; NFAT3;
KW   transcription factor; cytokine; gene expression; binding assay;
KW   immune system disease; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          397..686
FT                   /label= Rel_domain
XX
PN   WO9626959-A1.
XX
PD   06-SEP-1996.
XX
PF   04-MAR-1996;   96WO-US03113.
XX
PR   02-MAR-1995;   95US-0396479.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Hoey T;
XX
DR   WPI; 1996-412738/41.
DR   N-PSDB; AAT36868.
XX
PT   DNA mol. encoding human nuclear factors of activated T cells -
PT   useful for screening potential therapeutic and diagnostic agents for
PT   immune system diseases
XX
PS   Claim 5; Page 43-47; 64pp; English.
XX
CC   The amino acid sequence (AAW02250) of human nuclear factor of activated
CC   T-cells class 3, NAFT3, was deduced from an isolated cDNA clone
CC   (AAT36868).  NFATs (see also AAW02248-49 and AAW02251-53) include
CC   regulators of cytokine gene expression that modulate immune system
```

```
CC   function.  They have invariant rel domain peptides (see also
CC   AAW02254-55) and share at least 50% sequence identity in their rel
CC   domains.  Recombinant NFATs, or NFAT fragments contg. at least part
CC   of the rel domain, can be expressed in prokaryotic or eukaryotic host
CC   cells.  They are used in high-throughput screenings to identify
CC   agents useful in the diagnosis or treatment of diseases associated
CC   with expression of a gene modulated by a transcription complex contg.
CC   NFAT(s).
XX
SQ   Sequence   902 AA;

  Query Match             26.6%;  Score 122;  DB 17;  Length 902;
  Best Local Similarity   40.0%;  Pred. No. 2.1e-06;
  Matches   34;  Conservative   11;  Mismatches   34;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :      : || |||:|| :||  ||   |  | : | |    | ||    ||
Db     311 SPGPFDYVGAPPAESIPQKTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 368

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             :|: :|  :|:|||| |||  :
Db     369 ---RKEVAGMDYLAVPSPLAWSKAR 390


RESULT 4
AAU32449
ID   AAU32449 standard; Protein; 960 AA.
XX
AC   AAU32449;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #2940.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
XX
PS   Claim 20; Page 616; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon  and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence   960 AA;

  Query Match             26.6%;  Score 122;  DB 22;  Length 960;
  Best Local Similarity   40.0%;  Pred. No. 2.3e-06;
  Matches   34;  Conservative   11;  Mismatches   34;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           || || :      : || |||:|| :||  ||   |  | : | |    | ||    ||
Db     285 SPGPFDYVGAPPAESIPQKTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 342

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             :|: :|  :|:|||| |||  :
Db     343 ---RKEVAGMDYLAVPSPLAWSKAR 364


RESULT 5
AAW85730
ID   AAW85730 standard; Protein; 902 AA.
XX
AC   AAW85730;
XX
DT   27-SEP-1999  (first entry)
XX
DE   NF-AT3.
XX
KW   NF-AT3; hypertrophy; cardiomyocytes; cardiac hypertrophic response;
KW   heart failure; transgenic animals; screening; treatment; inhibition.
XX
OS   Mus musculus.
XX
PN   WO9919471-A1.
XX
PD   22-APR-1999.
XX
PF   15-OCT-1998;   98WO-US21845.
XX
PR   16-APR-1998;   98US-0061417.
PR   16-OCT-1997;   97US-0062864.
PR   10-NOV-1997;   97US-0065178.
PR   15-APR-1998;   98US-0081853.
XX
PA   (UYNT-) UNIV NORTH TEXAS HEALTH SCI CENT.
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Grant SR,  Molkentin JD,  Olson EN;
XX
DR   WPI; 1999-277635/23.
DR   N-PSDB; AAX08717.
XX
PT   Treating hypertrophy in cardiomyocytes by inhibiting NF-A3
XX
PS   Disclosure; Page 96-99; 105pp; English.
XX
CC   Hypertrophy in cardiomyocytes is treated by inhibiting function of
CC   NF-AT3.  Activation of NF-AT3 mediates the calcium ion-dependent
CC   cardiac hypertrophic response to a variety of stimuli, so inhibiting
CC   it can be used to treat or prevent cardiac hypertrophy and related
CC   heart failure. Transgenic animals, or cells, containing a
CC   constitutively active NF-AT3 gene can be used as models for screening
CC   modulators of hypertrophy and for studying human disease.
XX
SQ   Sequence   902 AA;

  Query Match             25.1%;  Score 115;  DB 20;  Length 902;
  Best Local Similarity   38.8%;  Pred. No. 1.7e-05;
  Matches   33;  Conservative   11;  Mismatches   35;  Indels    6;  Gaps    3;

Qy       2 SPAPFPFQYCVETD-IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
```

```
          || || :      : ||  ||:|| :||  ||   |  | : | |    | ||    ||
Db    311 SPGPFDYVGAPPAESIPQLTRRTSSEQAVALPRSEEPASCN-GKLPLGAEESVAPPGGS- 368

Qy     61 YPLKKDSSGDQFLSVPSPFTWSKPK 85
             :|: :|  :|:||||  ||| :
Db    369 ---RKEVAGMDYLAVPSPLAWSKAR 390


RESULT 6
AAR66874
ID   AAR66874 standard; Protein; 716 AA.
XX
AC   AAR66874;
XX
DT   12-AUG-1995  (first entry)
XX
DE   Human NF-AT120, C subfamily.
XX
KW   NF-AT120; nuclear factor of activated T cells 120 protein;
KW   transcription; differentiation; cytokine; T-cell; T-lymphocyte.
XX
OS   Homo sapiens.
XX
PN   WO9502053-A.
XX
PD   19-JAN-1995.
XX
PF   05-JUL-1994;   94WO-US07297.
XX
PR   06-JUL-1993;   93US-0088483.
PR   30-JUL-1993;   93US-0099998.
PR   30-AUG-1993;   93US-0113971.
PR   05-NOV-1993;   93US-0148061.
PR   04-APR-1994;   94US-0222626.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Arai N,  Masuda ES,  Tokumitsu H;
XX
DR   WPI; 1995-066898/09.
DR   N-PSDB; AAQ84301.
XX
PT   Nuclear Factor of Activated T cells 120 protein, antibody and
PT   nucleic acid - useful as transcriptional regulator of genes, such
PT   as cytokines
XX
PS   Disclosure; Page 63-67; 100pp; English.
XX
CC   A DNA fragment amplified from Jurkat cDNA by PCR had the sequence
CC   given in AAQ84300 and encoded NF-AT120 (Nuclear Factor of
CC   Activated T cells protein 120) (AAR66873).  Amplified DNA was used
CC   to screen for related NF-AT proteins by hybridization, leading to
CC   the discovery of 3 related subfamilies, designated class C, P and
CC   X (AAR66874, AAR66875, AAR75353).
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 16;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 7
AAR70357
ID   AAR70357 standard; Protein; 716 AA.
XX
AC   AAR70357;
XX
DT   22-NOV-1995  (first entry)
XX
DE   NF-AT transcription complex poypeptide.
XX
KW   NF-AT polypeptide; transcription complex; autoimmune diseases;
KW   immunomodulatory agents; T-cell related conditions;
KW   lymphocytic leukaemia; transplant rejection reactions.
XX
OS   Homo sapiens.
XX
PN   WO9508554-A.
XX
PD   30-MAR-1995.
XX
PF   20-SEP-1994;   94WO-US10724.
XX
PR   20-SEP-1993;   93US-0124981.
PR   13-JUN-1994;   94US-0260174.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Crabtree GR,  Ho SN,  Northrop JP;
XX
DR   WPI; 1995-139542/18.
XX
PT   DNA and polypeptide(s) encoding the transcription complex NF-AT
PT   - for use as immunomodulatory agents in the diagnosis and
PT   treatment of T-cell related conditions
XX
PS   Claim 11; Page 65; 84pp; English.
XX
CC   AAR70357 is a claimed human NF-AT transcription complex polypeptide,
CC   it can be used as an immunomodulatory agent, useful in the diagnosis
CC   and treatment of T-cell related conditions, e.g autoimmune diseases,
CC   lymphocytic leukaemias and transplant rejection reactions.
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 16;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 8
AAW02249
ID   AAW02249 standard; Protein; 716 AA.
XX
AC   AAW02249;
XX
DT   17-NOV-1996  (first entry)
XX
DE   Human transcription factor NFATc.
XX
KW   Nuclear factor of activated T-cells; NFAT; NFATc;
KW   transcription factor; cytokine; gene expression; binding assay;
KW   immune system disease; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          406..697
FT                   /label= Rel_domain
```

XX
PN WO9626959-A1.
XX
PD 06-SEP-1996.
XX
PF 04-MAR-1996; 96WO-US03113.
XX
PR 02-MAR-1995; 95US-0396479.
XX
PA (TULA-) TULARIK INC.
XX
PI Hoey T;
XX
DR WPI; 1996-412738/41.
DR N-PSDB; AAT36867.
XX
PT DNA mol. encoding human nuclear factors of activated T cells -
PT useful for screening potential therapeutic and diagnostic agents for
PT immune system diseases
XX
PS Claim 4; Page 41-43; 64pp; English.
XX
CC The amino acid sequence (AAW02249) of human nuclear factor of activated
CC T-cells class c, NAFTc, was deduced from an isolated cDNA clone
CC (AAT36867). NFATs (see also AAW02249 and AAW02250-53) include
CC regulators of cytokine gene expression that modulate immune system
CC function. They have invariant rel domain peptides (see also AAW02254-55)
CC and share at least 50% sequence identity in their rel domains.
CC Recombinant NFATs, or NFAT fragments contg. at least part of the rel
CC domain, can be expressed in prokaryotic or eukaryotic host cells.
CC They are used in high-throughput screenings to identify agents useful
CC in the diagnosis or treatment of diseases associated with expression
CC of a gene modulated by a transcription complex contg. NFAT(s).
XX
SQ Sequence 716 AA;

Query Match 24.8%; Score 114; DB 17; Length 716;
Best Local Similarity 34.7%; Pred. No. 1.6e-05;
Matches 25; Conservative 19; Mismatches 26; Indels 2; Gaps 2;

Qy 16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
:|:|:|||: :| : |:| :| |: | : : :| |: ::| ||:|:|
Db 331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy 76 PS-PFTWSKPKP 86
| |: |:||||
Db 390 PQHPYQWAKPKP 401

RESULT 9
AAY96535
ID AAY96535 standard; Protein; 716 AA.
XX
AC AAY96535;
XX
DT 12-SEP-2000 (first entry)
XX
DE Human nuclear factor of activated T cells c1 (NF-ATc1).
XX
KW NF-ATc1; cardiac hypertrophy; nuclear factor of activated T cells;
KW antagonist; congestive heart disease; cardiant.
XX
OS Homo sapiens.
XX
PN WO200030671-A2.
XX
PD 02-JUN-2000.
XX
PF 23-NOV-1999; 99WO-US27862.
XX
PR 24-NOV-1998; 98US-0198977.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Crabtree GR, Northrop JP, Ho SN;
XX
DR WPI; 2000-399929/34.
DR N-PSDB; AAA29252.
XX
PT Treating cardiac hypertrophy using NF-AT antagonists in
PT patients suffering from congestive heart disease
XX
PS Disclosure; Fig 1A-B; 139pp; English.
XX
CC A novel method for preventing and/or reducing cardiac hypertrophy in a
CC patient, comprises administering an NF-AT (nuclear factor of activated
CC T cells) antagonist to decrease the biological activity of NF-AT in
CC myocardial tissue (therefore preventing and/or reducing the level of
CC cardiac hypertrophy). The antagonists may decrease the transcriptional
CC activity, nuclear translocation or dephosphorylation of NF-AT, inhibit
CC binding of calcineurin to NF-AT, stimulate phosphorylation of NF-AT (e.g.
CC by increasing GSK-3) or inhibit formation of an NF-AT complex. The
CC antagonist is an antagonist of NF-ATc4 (also known as NF-AT3) and not
CC NF-ATc1, NF-ATc2 and NF-ATc3 (also known as NF-AT4). The method is used
CC for preventing and/or reducing cardiac hypertrophy in a patient suffering
CC from congestive heart disease (claimed) and for preventing other growth
CC of cardiac and vascular tissue.
XX
SQ Sequence 716 AA;

Query Match 24.8%; Score 114; DB 21; Length 716;
Best Local Similarity 34.7%; Pred. No. 1.6e-05;
Matches 25; Conservative 19; Mismatches 26; Indels 2; Gaps 2;

Qy 16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
:|:|:|||: :| : |:| :| |: | : : :| |: ::| ||:|:|
Db 331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy 76 PS-PFTWSKPKP 86
| |: |:||||
Db 390 PQHPYQWAKPKP 401

RESULT 10
AAY96565
ID AAY96565 standard; Protein; 716 AA.
XX
AC AAY96565;
XX
DT 12-SEP-2000 (first entry)
XX
DE Human nuclear factor of activated T cells c1 (NF-ATc1).
XX
KW NF-ATc1; cardiac hypertrophy; nuclear factor of activated T cells;
KW antagonist; congestive heart disease; cardiant.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 172..301
FT /note= "optionally contains at least one phosphoserine"
FT Domain 172..194
FT /label= SRR
FT /note= "NLS interacting domain"
FT Domain 199..219
FT /label= SP1
FT /note= "NLS interacting domain"
FT Domain 233..252
FT /label= SP2
FT /note= "NLS interacting domain"
FT Misc-difference 265..267
FT /label= Nuclear_localization_signal
FT /note= "N-terminal"
FT Domain 278..301
FT /label= SP3
FT /note= "NLS interacting domain"

```
FT   Misc-difference 681..685
FT                   /label= Nuclear_localization_signal
FT                   /note= "C-terminal"
XX
PN   WO200030671-A2.
XX
PD   02-JUN-2000.
XX
PF   23-NOV-1999;   99WO-US27862.
XX
PR   24-NOV-1998;   98US-0198977.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Crabtree GR,  Northrop JP,  Ho SN;
XX
DR   WPI; 2000-399929/34.
XX
PT   Treating cardiac hypertrophy using NF-AT antagonists in
PT   patients suffering from congestive heart disease
XX
PS   Disclosure; Page 49; 139pp; English.
XX
CC   A novel method for preventing and/or reducing cardiac hypertrophy in a
CC   patient, comprises administering an NF-AT (nuclear factor of activated
CC   T cells) antagonist to decrease the biological activity of NF-AT in
CC   myocardial tissue (therefore preventing and/or reducing the level of
CC   cardiac hypertrophy). The antagonists may decrease the transcriptional
CC   activity, nuclear translocation or dephosphorylation of NF-AT, inhibit
CC   binding of calcineurin to NF-AT, stimulate phosphorylation of NF-AT (e.g.
CC   by increasing GSK-3) or inhibit formation of an NF-AT complex. The
CC   antagonist is an antagonist of NF-ATc4 (also known as NF-AT3) and not
CC   NF-ATc1, NF-ATc2 and NF-ATc3 (also known as NF-AT4). The method is used
CC   for preventing and/or reducing cardiac hypertrophy in a patient suffering
CC   from congestive heart disease (claimed) and for preventing other growth
CC   of cardiac and vascular tissue.
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 21;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 11
AAB90774
ID   AAB90774 standard; Protein; 716 AA.
XX
AC   AAB90774;
XX
DT   15-JUN-2001  (first entry)
XX
DE   Human shear stress-response protein SEQ ID NO: 48.
XX
KW   Human; shear stress-response protein; vascular disease;
KW   arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200125427-A1.
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-JP06840.
XX
PR   01-OCT-1999;   99JP-0280976.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (NOJI/) NOJIMA H.
XX
PI   Nojima H,  Yoshisue H,  Obayashi M,  Ota T,  Kawabata A,  Sakurada K;
PI   Kuga T,  Sekine S,  Nakamura Y,  Sugano S;
XX
DR   WPI; 2001-266308/27.
DR   N-PSDB; AAH02897.
XX
PT   DNA sequences, proteins encoded by them and antibodies against them
PT   useful in diagnosis and treatment of vascular disease caused by
PT   arteriosclerosis -
XX
PS   Claim 60; Page 352-356; 678pp; Japanese.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human shear stress response proteins. These are useful in the
CC   diagnosis, treatment and screening of vascular diseases caused by
CC   arteriosclerosis, including heart failure, post-PTCA restenosis and
CC   hypertension.
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 22;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 12
AAB66497
ID   AAB66497 standard; Protein; 716 AA.
XX
AC   AAB66497;
XX
DT   10-APR-2001  (first entry)
XX
DE   Human NF-AT protein.
XX
KW   Human; nuclear factor of activated T lymphocytes; NF-AT;
KW   NF-AT cytoplasmic component; NF-ATc; NF-AT translocation;
KW   nuclear localisation sequence; NLS; SRR.
XX
OS   Homo sapiens.
XX
PN   US6171781-B1.
XX
PD   09-JAN-2001.
XX
PF   27-MAR-1998;   98US-0049691.
XX
PR   20-SEP-1993;   93US-0124981.
PR   13-JUN-1994;   94US-0260174.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Crabtree GR,  Northrop JP,  Ho SN;
XX
DR   WPI; 2001-122328/13.
XX
PT   Screening assay for identifying modulators of translocation of nuclear
PT   factor of activated T lymphocytes across nuclear membrane of cell, by
PT   treating with nuclear factor of activated T lymphocyte polypeptide  -
XX
```

```
PS   Claim 10; Column 40; 99pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   identifying a compound which modulates translocation of a nuclear factor
CC   of activated T lymphocytes (NF-AT) polypeptide across the nuclear
CC   membrane of a cell. The method involves binding the compound to the
CC   NF-AT polypeptide. The method is useful for identifying compounds which
CC   modulate nuclear translocation of NF-AT present in the cytoplasm and/or
CC   in the nucleus of the cell and for identifying agents that modulate
CC   phosphorylation/dephosphorylation of NF-AT.
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 22;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 13
AAB66501
ID   AAB66501 standard; Protein; 716 AA.
XX
AC   AAB66501;
XX
DT   10-APR-2001  (first entry)
XX
DE   Human NF-ATc protein.
XX
KW   Human; nuclear factor of activated T lymphocytes; NF-AT;
KW   NF-AT cytoplasmic component; NF-ATc; NF-AT translocation;
KW   nuclear localisation sequence; NLS; SRR.
XX
OS   Homo sapiens.
XX
PN   US6171781-B1.
XX
PD   09-JAN-2001.
XX
PF   27-MAR-1998;   98US-0049691.
XX
PR   20-SEP-1993;   93US-0124981.
PR   13-JUN-1994;   94US-0260174.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Crabtree GR,  Northrop JP,  Ho SN;
XX
DR   WPI; 2001-122328/13.
DR   N-PSDB; AAF31710.
XX
PT   Screening assay for identifying modulators of translocation of nuclear
PT   factor of activated T lymphocytes across nuclear membrane of cell, by
PT   treating with nuclear factor of activated T lymphocyte polypeptide -
XX
PS   Disclosure; Fig 1; 99pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   identifying a compound which modulates translocation of a nuclear factor
CC   of activated T lymphocytes (NF-AT) polypeptide across the nuclear
CC   membrane of a cell. The method involves binding the compound to the
CC   NF-AT polypeptide. The method is useful for identifying compounds which
CC   modulate nuclear translocation of NF-AT present in the cytoplasm and/or
CC   in the nucleus of the cell and for identifying agents that modulate
CC   phosphorylation/dephosphorylation of NF-AT.
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 22;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQQPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 14
AAE21559
ID   AAE21559 standard; Protein; 716 AA.
XX
AC   AAE21559;
XX
DT   16-JUL-2002  (first entry)
XX
DE   Human cytoplasmic nuclear factor of activated T cell (NF-ATc) protein #1.
XX
KW   Human; immunosuppressive; cytoplasmic nuclear factor of activated T cell;
KW   NF-ATc; nuclear translocation.
XX
OS   Homo sapiens.
XX
PN   US6352830-B1.
XX
PD   05-MAR-2002.
XX
PF   15-JAN-1999;   99US-0232346.
XX
PR   22-AUG-1991;   91US-0749385.
PR   18-APR-1994;   94US-0228944.
PR   20-SEP-1993;   93US-0124981.
PR   13-JUN-1994;   94US-0260174.
PR   31-JUL-1995;   95US-0507032.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Crabtree GR,  Northrop JP,  Ho SN,  Flanagan WM;
XX
DR   WPI; 2002-314700/35.
XX
PT   Identifying immunosuppressive agent comprises contacting cell having
PT   cytoplasmic NF-AT polypeptide with inducer of polypeptide cytoplasmic
PT   translocation, in presence and absence of test agent, and assaying the
PT   translocation -
XX
PS   Example 9; Column 67-72; 83pp; English.
XX
CC   The invention relates to a method for identifying an immunosuppressive
CC   agent. The method comprising: contacting a cell containing cytoplasmic
CC   nuclear factor of activated T cell (NF-ATc) polypeptide with a compound
CC   that induces nuclear translocation of the polypeptide; and nuclear
CC   translocation of the NF-ATc is assayed. The method is useful for
CC   identifying an immunosuppressive agent and an immune regulating agent.
CC   The present sequence is human NF-ATc protein.
XX
SQ   Sequence   716 AA;

  Query Match             24.8%;  Score 114;  DB 23;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389
```

```
Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


RESULT 15
AAE21563
ID   AAE21563 standard; Protein; 716 AA.
XX
AC   AAE21563;
XX
DT   16-JUL-2002  (first entry)
XX
DE   Human cytoplasmic nuclear factor of activated T cell (NF-ATc) protein #2.
XX
KW   Human; immunosuppressive; cytoplasmic nuclear factor of activated T cell;
KW   NF-ATc; nuclear translocation.
XX
OS   Homo sapiens.
XX
PN   US6352830-B1.
XX
PD   05-MAR-2002.
XX
PF   15-JAN-1999;   99US-0232346.
XX
PR   22-AUG-1991;   91US-0749385.
PR   18-APR-1994;   94US-0228944.
PR   20-SEP-1993;   93US-0124981.
PR   13-JUN-1994;   94US-0260174.
PR   31-JUL-1995;   95US-0507032.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Crabtree GR,  Northrop JP,  Ho SN,  Flanagan WM;
XX
DR   WPI; 2002-314700/35.
DR   N-PSDB; AAD34036.
XX
PT   Identifying immunosuppressive agent comprises contacting cell having
PT   cytoplasmic NF-AT polypeptide with inducer of polypeptide cytoplasmic
PT   translocation, in presence and absence of test agent, and assaying the
PT   translocation  -
XX
PS   Claim 20; Column 81-84; 83pp; English.
XX
CC   The invention relates to a method for identifying an immunosuppressive
CC   agent. The method comprising: contacting a cell containing cytoplasmic
CC   nuclear factor of activated T cell (NF-ATc) polypeptide with a compound
CC   that induces nuclear translocation of the polypeptide; and nuclear
CC   translocation of the NF-ATc is assayed. The method is useful for
CC   identifying an immunosuppressive agent and an immune regulating agent.
CC   The present sequence is human NF-ATc protein.
XX
SQ   Sequence   716 AA;

  Query Match            24.8%;  Score 114;  DB 23;  Length 716;
  Best Local Similarity  34.7%;  Pred. No. 1.6e-05;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy      16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
           :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db     331 VPVKSRKTTLEQQPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy      76 PS-PFTWSKPKP 86
           |  |: |:||||
Db     390 PQHPYQWAKPKP 401


Search completed: April 11, 2003, 14:22:23
Job time : 40 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 14:21:02 ; Search time 17 Seconds
(without alignments)
486.327 Million cell updates/sec

Title: US-09-550-115A-2
Perfect score: 459
Sequence: 1 LSPAPFPFQYCVETDIPLKT.....SSGDQFLSVPSPFTWSKPKP 86

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	459	100.0	1065	2	A57410	transcription fact
2	430	93.7	1075	2	A57377	transcription fact
3	114	24.8	716	2	S45262	NF-AT component -
4	96.5	21.0	718	2	JC5805	transcription fact
5	81.5	17.8	329	2	G86253	hypothetical prote
6	73.5	16.0	3488	2	T34418	hypothetical prote
7	72	15.7	2130	2	AB0821	probable exported
8	71.5	15.6	792	2	JC7122	protein kinase (EC
9	70	15.3	643	2	S54395	permease-like prot
10	70	15.3	649	1	B37953	transcription regu
11	70	15.3	652	2	G95177	hypothetical prote
12	70	15.3	652	2	E98044	hypothetical prote
13	70	15.3	660	2	G97912	hypothetical prote
14	70	15.3	661	2	E95042	hypothetical prote
15	69.5	15.1	638	2	B35816	transcription regu
16	69	15.0	649	2	A35816	transcription regu
17	67.5	14.7	314	2	I38864	E1B 19K/Bcl-2-inte
18	67	14.6	177	2	H90631	probable fimbrial
19	67	14.6	177	2	G85482	probable type-1 fi
20	67	14.6	1112	2	T32733	AMPA glutamate rec
21	66.5	14.5	1039	2	T28905	hypothetical prote
22	66	14.4	728	1	A60185	hepatocyte growth
23	66	14.4	1944	2	A59438	KIAA1424 protein [
24	65.5	14.3	652	2	S47979	aliB protein precu
25	65	14.2	635	2	T20587	hypothetical prote
26	64.5	14.1	421	2	I49734	HNF-3/fork-head ho
27	64	13.9	84	2	S76062	hypothetical prote
28	64	13.9	616	2	S38060	carboxylic acid tr
29	64	13.9	728	1	A35644	hepatocyte growth
30	63	13.7	498	2	T09436	gag polyprotein -
31	63	13.7	520	2	S47142	mating type A prot
32	63	13.7	570	2	T49181	cyclophylin-like p
33	63	13.7	1216	2	AH1535	pyruvate-flavodoxi
34	62.5	13.6	474	2	T39587	serine/threonine-p
35	62.5	13.6	671	2	S61693	probable membrane
36	62.5	13.6	1530	2	E82085	glutamate synthase
37	62	13.5	294	2	G97484	dihydrodipicolinat
38	62	13.5	294	2	AG2702	dihydrodipicolinat
39	62	13.5	579	2	AF2094	hypothetical prote
40	62	13.5	963	2	T40290	hypothetical prote
41	62	13.5	1127	2	T30334	AND-1 protein - Af
42	62	13.5	1856	2	C95008	immunoglobulin A1
43	61.5	13.4	279	2	C75491	probable serine es
44	61.5	13.4	441	2	B86252	hypothetical prote
45	61.5	13.4	838	2	A96557	probable receptor

ALIGNMENTS

RESULT 1
A57410
transcription factor NFATc3 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 08-Feb-1996
C;Accession: A57410
R;Ho, S.N.; Thomas, D.J.; Timmerman, L.A.; Li, X.; Francke, U.; Crabtree, G.R.
J. Biol. Chem. 270, 19898-19907, 1995
A;Title: NFATc3, a lymphoid-specific NFATc family member that is calcium-regulated an
A;Reference number: A57410; MUID:95378239; PMID:7650004
A;Accession: A57410
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1065 <HOA>
A;Cross-references: GB:U28807

Query Match 100.0%; Score 459; DB 2; Length 1065;
Best Local Similarity 100.0%; Pred. No. 5.8e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     311 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 370

Qy      61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
           ||||||||||||||||||||||||||
Db     371 YPLKKDSSGDQFLSVPSPFTWSKPKP 396
```

RESULT 2
A57377
transcription factor NFATx - human
C;Species: Homo sapiens (man)
C;Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 05-Nov-1999
C;Accession: A57377
R;Masuda, E.S.; Naito, Y.; Tokumitsu, H.; Campbell, D.; Saito, F.; Hannum, C.; Arai,
Mol. Cell. Biol. 15, 2697-2706, 1995
A;Title: NFATx, a novel member of the nuclear factor of activated T cells family that
A;Reference number: A57377; MUID:95257951; PMID:7739550
A;Accession: A57377
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1075 <MAS>
A;Cross-references: GB:U14510; NID:g780373; PIDN:AAA86308.1; PID:g780374
C;Keywords: transcription factor

Query Match 93.7%; Score 430; DB 2; Length 1075;
Best Local Similarity 91.9%; Pred. No. 8.7e-39;
Matches 79; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 LSPAPFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQ 60
           | || ||||||||||||||||||||||||||||||:||||||:|||:||||:|||||||
```

```
Db    321 LGPAVFPFQYCVETDIPLKTRKTSEDQAAILPGKLELCSDDQGSLSPARETSIDDGLGSQ 380

Qy     61 YPLKKDSSGDQFLSVPSPFTWSKPKP 86
          ||||||| ||||||||||||||||||
Db    381 YPLKKDSCGDQFLSVPSPFTWSKPKP 406


RESULT 3
S45262
NF-AT component - human
C;Species: Homo sapiens (man)
C;Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C;Accession: S45262
R;Northrop, J.P.; Ho, S.N.; Chen, L.; Thomas, D.J.; Timmerman, L.A.; Nolan, G.P.; Admon,
Nature 369, 497-502, 1994
A;Title: NF-AT components define a family of transcription factors targeted in T-cell ac
A;Reference number: S45262; MUID:94261186; PMID:8202141
A;Accession: S45262
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-716 <NOR>
A;Cross-references: GB:U08015; NID:g500631; PIDN:AAA19601.1; PID:g500632

  Query Match             24.8%;  Score 114;  DB 2;  Length 716;
  Best Local Similarity   34.7%;  Pred. No. 0.00021;
  Matches   25;  Conservative   19;  Mismatches   26;  Indels    2;  Gaps    2;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFLSV 75
          :|:|:|||: :|   :  |:|   :| |:  |  : : :|    |: ::|    ||:|:|
Db    331 VPVKSRKTTLEQPPSVALKVEPVGEDLGSPPPPADFAPEDYSSFQH-IRKGGFCDQYLAV 389

Qy     76 PS-PFTWSKPKP 86
          |  |: |:||||
Db    390 PQHPYQWAKPKP 401


RESULT 4
JC5805
transcription factor NFATc - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 07-May-1999
C;Accession: JC5805
R;Pan, S.; Koyano-Nakagawa, N.; Tsuruta, L.; Amasaki, Y.; Yokota, T.; Mori, S.; Arai, N.
Biochem. Biophys. Res. Commun. 240, 314-323, 1997
A;Title: Molecular cloning and functional characterization of murine cDNA encoding trans
A;Reference number: JC5805; MUID:98049829; PMID:9388475
A;Accession: JC5805
A;Molecule type: mRNA
A;Residues: 1-718 <PAN>
A;Note: the sequences of residues 30-39 and 40-59 are interchanged in the authors' trans
C;Comment: This protein plays a role in immune and inflammatory response by regulating t
F;202-211,236-245,281-290/Region: SP-box
F;684-687/Region: nuclear location signal

  Query Match             21.0%;  Score 96.5;  DB 2;  Length 718;
  Best Local Similarity   30.4%;  Pred. No. 0.017;
  Matches   24;  Conservative   17;  Mismatches   19;  Indels   19;  Gaps    3;

Qy     16 IPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSVDDGLGSQYP--------LKKDS 67
          :|:|:|||: : |  :  |:|   :| |   |:          | :|        |:| :
Db    334 VPIKSRKTALEHAPSVALKVEPAGEDLGTTPPT----------SDFPPEEYTFQHLRKGA 383

Qy     68 SGDQFLSVP-SPFTWSKPK 85
          :|:|||| : : |:|||
Db    384 FCEQYLSVPQASYQWAKPK 402


RESULT 5
G86253
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86253
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86253
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-329 <STO>
A;Cross-references: GB:AE005172; NID:g3157936; PIDN:AAC17619.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

  Query Match             17.8%;  Score 81.5;  DB 2;  Length 329;
  Best Local Similarity   25.3%;  Pred. No. 0.3;
  Matches   20;  Conservative   10;  Mismatches   26;  Indels   23;  Gaps    1;

Qy     27 QAAILPGKLEICSDDQGNLSPSRETSVDDGLGS----------------------QYPL 63
          |::    :  :|    ||:  :   |:||| ||                      :||:
Db     20 QSSTTSSQSNLCRSSCGNIPINYPFSIDDGCGSPYYRHMLICSDNDTKLELRTPSGKYPV 79

Qy     64 KKDSSGDQFLSVPSPFTWS 82
          |  |  |  | |  || |:
Db     80 KSISYSDPHLLVSDPFMWN 98


RESULT 6
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

  Query Match             16.0%;  Score 73.5;  DB 2;  Length 3488;
  Best Local Similarity   34.4%;  Pred. No. 36;
  Matches   22;  Conservative    9;  Mismatches   30;  Indels    3;  Gaps    2;

Qy     15 DIPLKTRKTSEDQAAI-LPGKLEICSDDQGNLSPSRETSVDDGLGSQYPLKKDSSGDQFL 73
          |:  |::||:|||. :    |||  :|    :    || |||    :   ||    | |:
Db    697 DVTDKSKKTTEDQTKVATDSKLEKAADTTKQI--ETETVVDDKSKKKVLKKKTEKSDSFI 754

Qy     74 SVPS 77
          |  |
Db    755 SQKS 758


RESULT 7
AB0821
probable exported protein STY2760 [imported] - Salmonella enterica subsp. enterica se
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AB0821
```

```
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; PMID:11677608
A;Accession: AB0821
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2130 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02718.1; PID:g16503731; GSPDB:GN00176
C;Genetics:
A;Gene: STY2760

  Query Match             15.7%;  Score 72;  DB 2;  Length 2130;
  Best Local Similarity   27.0%;  Pred. No. 30;
  Matches   24;  Conservative   13;  Mismatches   38;  Indels   14;  Gaps    2;

Qy       5 PFPFQYCVETDIPLKTRKTSEDQAAILPGKLEICSDDQGNLSPSRETSV---------- 53
           | |::|      | | ||  | | |    :: : |:|    ||   | |
Db     501 PLPYRYMELFIDPAKNRK-GEHQDAWDNQRVTVSSEDMRASSPEHYTGVTDVNGQAHLTL 559

Qy      54 --DDGLGSQYPLKKDSSGDQFLSVPSPFT 80
             | |:| : |::     |:  :|  ||:
Db     560 QHDSGMGGETPIRIVMPDDEGGNVELPFS 588


RESULT 8
JC7122
protein kinase (EC 2.7.1.37) 2 - mouse (strain balb/c)
N;Alternate names: serine (threonine) protein kinase
C;Species: Mus musculus (house mouse)
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C;Accession: JC7122
R;Hayashi, K.; Igarashi, H.; Ogawa, M.; Sakaguchi, N.
Biochem. Biophys. Res. Commun. 264, 449-456, 1999
A;Title: Activity and substrate specificity of the murine STK2 serine/threonine kinase t
A;Reference number: JC7122; MUID:20001940; PMID:10529384
A;Accession: JC7122
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-792 <HAY>
A;Cross-references: GB:AJ223071; NID:g4138208; PID:g4138209
C;Genetics:
A;Gene: MSTK2L
C;Superfamily: human serine/threonine-specific protein kinase STK2; protein kinase homol
C;Keywords: phosphotransferase

  Query Match             15.6%;  Score 71.5;  DB 2;  Length 792;
  Best Local Similarity   44.7%;  Pred. No. 11;
  Matches   17;  Conservative    2;  Mismatches   14;  Indels    5;  Gaps    1;

Qy      39 SDDQGNLSPSRETSVDDGLGSQ-----YPLKKDSSGDQ 71
           ||  ||| | | :   || ||:     ||  ||   ||
Db     418 SDQPGNLLPRRSSDGGDGEGSELVKPLYPSNKDQKPDQ 455


RESULT 9
S54395
permease-like protein - Streptococcus pneumoniae (fragment)
C;Species: Streptococcus pneumoniae
C;Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C;Accession: S54395
R;Pearce, B.J.; Naughton, A.M.; Masure, H.R.
Mol. Microbiol. 12, 881-892, 1994
A;Title: Peptide permeases modulate transformation in Streptococcus pneumoniae.
A;Reference number: S54395; MUID:95020610; PMID:7523829
A;Accession: S54395
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-643 <PEA>
A;Cross-references: EMBL:L20556; NID:g516639; PIDN:AAA26952.1; PID:g516640
C;Genetics:
A;Gene: plpA

  Query Match             15.3%;  Score 70;  DB 2;  Length 643;
  Best Local Similarity   34.1%;  Pred. No. 12;
  Matches   28;  Conservative   12;  Mismatches   26;  Indels   16;  Gaps    6;

Qy       8 FQYCVETDIP----LKTRK--TSEDQAAILPGKLEICSDDQGNLSPS--RETSVD-DGLG 58
           | |  |||      | | |  |:   : :: | ||  :|  ||  ||   : ||  |||
Db      19 FSYIYETDPDNLNYLTTAKAATANITSNVVDGLLE--NDRYGNFVPSMAEDWSVSKDGLT 76

Qy      59 SQYPLKKD-----SSGDQFLSV 75
             | ::||     | |::: :|
Db      77 YTYTIRKDAKWYTSEGEEYAAV 98


RESULT 10
B37953
transcription regulator PAN-2 - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: B37953
R;German, M.S.; Blanar, M.A.; Nelson, C.; Moss, L.G.; Rutter, W.J.
Mol. Endocrinol. 5, 292-299, 1991
A;Title: Two related helix-loop-helix proteins participate in separate cell-specific
A;Reference number: A37953; MUID:91246228; PMID:1710033
A;Accession: B37953
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-649 <GER>
C;Superfamily: human transcription factor 3

  Query Match             15.3%;  Score 70;  DB 1;  Length 649;
  Best Local Similarity   26.1%;  Pred. No. 12;
  Matches   30;  Conservative    6;  Mismatches   17;  Indels   62;  Gaps    6;

Qy      27 QAAILPGKLEICSDDQGNLSPS----------------RETSVDDGLGSQ---------- 60
           ||  |||:| : |   | ||||                |  : | || :|
Db     123 QAGFLPGELGLSS--PGPLSPSGVKSGSQYYPSYPSNPRRRAADSGLDTQSKKVRKVPPG 180

Qy      61 -----YPLKKDSSGDQF-------------LSVPSPF-----------TWSKP 84
                ||    |||| :               : ||||             || |
Db     181 LPSSVYP---SSSGDSYGRDAAAYPSAKTPGSAYPSPFYVADGSLHPSAELWSPP 232


RESULT 11
G95177
hypothetical protein SP1527 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: G95177
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: G95177
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75616.1; PID:g14973017; GSPDB:GN00164; TIGR:
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1527

  Query Match             15.3%;  Score 70;  DB 2;  Length 652;
  Best Local Similarity   30.3%;  Pred. No. 12;
  Matches   27;  Conservative   13;  Mismatches   31;  Indels   18;  Gaps    5;

Qy       2 SPAPFPFQYCVETDIPLKTRKTSEDQAA-------ILPGKLEICSDDQGNLSPSRE---T 51
           | |   : |   :| | |      :|::||       :: | ||  :|  ||: ||     |
```

```
Db      29 STASKTYNYVYSSD-PSSLNYLAENRAATSDIVANLVDGLLE--NDQYGNIIPSLAEDWT 85

Qy      52 SVDDGLGSQYPLKKD-----SSGDQFLSV 75
           |||   | |:||     | |:::  |
Db      86 VSQDGLTYTYKLRKDAKWFTSEGEEYAPV 114


RESULT 12
E98044
hypothetical protein aliB [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: E98044
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E98044
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00186.1; PID:g15459032; GSPDB:GN00174
C;Genetics:
A;Gene: aliB

  Query Match             15.3%;  Score 70;  DB 2;  Length 652;
  Best Local Similarity   30.3%;  Pred. No. 12;
  Matches   27;  Conservative   13;  Mismatches   31;  Indels   18;  Gaps    5;

Qy       2 SPAPFPFQYCVETDIPLKTRKTSEDQAA-------ILPGKLEICSDDQGNLSPSRE---T 51
           | |   : |   :| |      :|::||       :: | ||  :|  ||: ||     |
Db     -29 STASKTYNYVYSSD-PSSLNYLAENRAATSDIVANLVDGLLE--NDQYGNIIPSLAEDWT 85

Qy      52 SVDDGLGSQYPLKKD-----SSGDQFLSV 75
           |||   | |:||     | |:::  |
Db      86 VSQDGLTYTYKLRKDAKWFTSEGEEYAPV 114


RESULT 13
G97912
hypothetical protein aliA [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: G97912
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: G97912
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-660 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99131.1; PID:g15457883; GSPDB:GN00174
C;Genetics:
A;Gene: aliA

  Query Match             15.3%;  Score 70;  DB 2;  Length 660;
  Best Local Similarity   34.1%;  Pred. No. 12;
  Matches   28;  Conservative   12;  Mismatches   26;  Indels   16;  Gaps    6;

Qy       8 FQYCVETDIP----LKTRK--TSEDQAAILPGKLEICSDDQGNLSPS--RETSVD-DGLG 58
           | |  |||      | | |  |:   : :: | ||  :|  ||  ||   : ||  |||
Db      36 FSYIYETDPDNLNYLTTAKAATANITSNVVDGLLE--NDRYGNFVPSMAEDWSVSKDGLT 93

Qy      59 SQYPLKKD-----SSGDQFLSV 75
             | ::||     | |::: :|
Db      94 YTYTIRKDAKWYTSEGEEYAAV 115


RESULT 14
E95042
hypothetical protein SP0366 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: E95042
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: E95042
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-661 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK74534.1; PID:g14971835; GSPDB:GN00164; TIGR:
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0366

  Query Match             15.3%;  Score 70;  DB 2;  Length 661;
  Best Local Similarity   34.1%;  Pred. No. 12;
  Matches   28;  Conservative   12;  Mismatches   26;  Indels   16;  Gaps    6;

Qy       8 FQYCVETDIP----LKTRK--TSEDQAAILPGKLEICSDDQGNLSPS--RETSVD-DGLG 58
           | |  |||      | | |  |:   : :: | ||  :|  ||  ||   : ||  |||
Db      37 FSYIYETDPDNLNYLTTAKAATANITSNVVDGLLE--NDRYGNFVPSMAEDWSVSKDGLT 94

Qy      59 SQYPLKKD-----SSGDQFLSV 75
             | ::||     | |::: :|
Db      95 YTYTIRKDAKWYTSEGEEYAAV 116


RESULT 15
B35816
transcription regulator Pan-1 - rat
N;Alternate names: insulin enhancer-binding protein 1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 16-Jul-1999
C;Accession: B35816; A35487
R;Nelson, C.; Shen, L.P.; Meister, A.; Fodor, E.; Rutter, W.J.
Genes Dev. 4, 1035-1043, 1990
A;Title: Pan: a transcriptional regulator that binds chymotrypsin, insulin, and AP-4
A;Reference number: A35816; MUID:90346284; PMID:2200736
A;Accession: B35816
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-638 <NEL>
A;Cross-references: GB:X62323; NID:g35277; PIDN:CAA44199.1; PID:g1045195
R;Shibasaki, Y.; Sakura, H.; Takaku, F.; Kasuga, M.
Biochem. Biophys. Res. Commun. 170, 314-321, 1990
A;Title: Insulin enhancer binding protein has helix-loop-helix structure.
A;Reference number: A35487; MUID:90321246; PMID:2196879
A;Accession: A35487
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 297-377,'Q',378-638 <SHI>
C;Superfamily: human transcription factor 3
C;Keywords: alternative splicing; DNA binding; nucleus; transcription regulation

  Query Match             15.1%;  Score 69.5;  DB 2;  Length 638;
  Best Local Similarity   25.7%;  Pred. No. 14;
  Matches   29;  Conservative    5;  Mismatches   22;  Indels   57;  Gaps    5;

Qy      27 QAAILPGKLEICSDDQGNLSPS---------------RETSVDDGLGSQYPLK------ 64
           ||  |||:| : |   | ||||               |  : | ||    | |
Db     114 QAGFLPGELGLSS--PGPLSPSGVKSSSQYYTSFPSNPRRRAADGGLADTQPKKVRKVPP 171

Qy      65 -------KDSSGDQF-------------LSVPSPF-----------TWSKP 84
```

```
                    ||||  :                :  ||||             ||  |
Db    172 GLPSSVYPSSSGDNYSRDATAYPSAKTPSSAYPSPFYVADGSLHPSAELWSPP 224
```

Search completed: April 11, 2003, 14:23:42
Job time : 21 secs

WEST Search History

DATE: Monday, April 21, 2003

Set Name side by side	Query	Hit Count	Set Name result set
DB=USPT,PGPB,JPAB,EPAB; PLUR=YES; OP=OR			
L17	L16 and (l8 or l9 or l10)	19	L17
L16	L15 and l2	258	L16
L15	L14 and l1	1224	L15
L14	fk506 or fk-506 or "fk 506"	2736	L14
L13	fk506 or fk-506 or :fk 506	101057	L13
L12	L11 and (l8 or l9 or l10)	24	L12
L11	l1 and l2	368	L11
L10	(((435/21)!.CCLS.))	625	L10
L9	(((435/4)!.CCLS.))	4330	L9
L8	((530/324)!.CCLS.)	3223	L8
L7	"calcineurin binding"	54	L7
L6	calcineurin.ti.	19	L6
L5	(l3 or l4) and l1	3	L5
L4	"ARAI; KEN-ICHI"	28	L4
L3	"LIU; JIE"	40	L3
L2	nfat$ or "nf at" or nf-at$	644	L2
L1	calcineurin or cn	88767	L1

END OF SEARCH HISTORY